Using Resource Depletion to Examine Processing of
Familiar and Unfamiliar Metaphors

By
FELIX S. PAMBUCCIAN
B.S., University of Minnesota, Minneapolis, 2013

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Chicago, 2017

Chicago, Illinois

Defense Committee:

Gary Raney, Chair and Advisor
Kara Morgan-Short
Susan Goldman

This thesis is dedicated to my parents, who have given me endless love and support throughout my academic career, and throughout my life.

ACKNOWLEDGMENTS

I would like to thank my thesis committee – Gary, Susan, and Kara – whose insightful thoughts and criticisms led me to ask new questions and hone my approach to this study. The finished product is far better for your input and guidance.

I would like to give special thanks to my advisor, Gary, whose support, guidance, and patience have been invaluable to me throughout the course of this project, and throughout my career as a graduate student.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

A study of the processes involved in metaphor comprehension was carried out using an experimental manipulation. Specifically, the predictions of two established models of metaphor comprehension were tested and compared. The central research question was whether familiar and unfamiliar metaphors use the same set of controlled comprehension processes, as specified by the Categorization model, or whether separate processes are recruited: an automatic process for comprehension of familiar metaphors and a controlled process for comprehension of unfamiliar metaphors, as specified by the Career of Metaphor model. A total of 160 participants, 80 in each experiment, were recruited for participation in two complementary experiments. In both experiments, participants were first exposed to a resource depletion manipulation in which resource depletion was induced in the experimental group but not the control group. Following the depletion manipulation, participants performed a subsequent task in which they read a set of metaphors and literal sentences and either indicated the moment they had understood each sentence (Experiment 1) or categorized each sentence as abstract or concrete in meaning (Experiment 2). Response times for the control and experimental groups in each experiment were recorded for comparison.

In Experiment 1, participants in the experimental condition were significantly slower than controls in comprehending low-familiarity and mid-familiarity metaphors, but not high-familiarity metaphors or literal sentences. This may indicate that comprehension of mid- and low-familiarity metaphors requires higher-order controlled processing, which was impaired in the experimental group following the depletion manipulation. In Experiment 2, no differences in response time were found between the control and experimental groups.

Overall, results support elements of both the Categorization and Career of Metaphor models, but support appears to be stronger for the Career of Metaphor model. The effect of

resource depletion on comprehension of metaphors varied based on which stimulus group (high-, mid-, and low-familiarity) the metaphor belonged to. Comprehension of low- and mid-familiarity metaphors was greatly affected by depletion, whereas comprehension of high-familiarity metaphors was only slightly affected. This may indicate that a separate process underlies comprehension of highly-familiar metaphors, consistent with the Career of Metaphor model.

I. INTRODUCTION

A. **Metaphor Processing**

Broadly speaking, a metaphor is a statement comparing two things which are not obviously similar (Gentner & Bowdle, 2002). A metaphor implies an underlying connection between the two items being compared, the nature of which must be understood in order to arrive at the intended meaning of the metaphor. Although traditionally more associated with poetry and literature than with everyday communication, metaphors are commonly used in everyday speech, with many abstract concepts, such as time, described in heavily metaphorical fashion (Lakoff & Johnson, 1980).

The basic structure of a metaphor phrase is *an X is a Y,* in which the *X* is known as the *target* word and *Y* is known as the *vehicle* word. Consider the metaphor phrase *love is a flower*, in which *love* is the target word and *flower* is the vehicle word. The literal interpretation of this phrase is nonsensical: love is *not* literally a flower. In order to arrive at a meaningful interpretation of this phrase, the figurative meaning of the vehicle word *flower* must be activated. In this case, one possible characteristic shared by both *flower* and *love* is that both are beautiful, so a possible figurative interpretation of *love is a flower* is, roughly, that love is beautiful.

Several theorists have attempted to describe the way in which people process and understand metaphors, focusing in particular on the way in which the figurative meaning of the metaphor phrase is activated and selected over the literal meaning. The first major attempt to describe metaphor comprehension can be found in the pragmatic theory of figurative language comprehension (Searle, 1975), also known as the Standard Pragmatic Model (SPM). The pragmatic theory posits a multi-stage view of metaphor processing. According to the pragmatic theory of figurative language, when reading or listening to a metaphor phrase, the reader/listener will first extract the literal meaning of the phrase, followed by a second stage in which they evaluate whether the literal meaning is the reasonable or intended meaning. If the reader/listener determines that the literal meaning is not plausible or not intended, they will proceed to a third stage in which they determine the intended (figurative) meaning based on the content of the metaphor phrase and the preceding context. According to the pragmatic theory, then, the literal meaning of a metaphor phrase is accessed and considered earlier in the comprehension process than the figurative meaning. Indeed, the figurative meaning is considered only if the literal interpretation has first been discarded as implausible; when the literal meaning is plausible, the comprehension process will not proceed past stage 2, and thus a figurative meaning will not be accessed or considered at all.

Although the idea that metaphors are initially treated as potential literal statements has some intuitive appeal, a series of later studies (e.g., Glucksberg, Gildea, & Bookin, 1982; Gibbs & Gerrig, 1989; Gildea & Glucksberg, 1983) have provided overwhelming evidence against this model. Glucksberg and colleagues (1982) examined the implications of the standard pragmatic theory, in particular the idea that a figurative meaning is only considered if the literal meaning is determined to be implausible. Glucksberg and colleagues presented subjects with sentences, some of which were metaphors, and asked them to respond true or false to each sentence based on whether or not the sentence was *literally* true. When a sentence was literally false, but a metaphoric interpretation of the sentence was possible, subjects responded slower compared to literally false sentences that had no plausible metaphoric interpretation. This result suggested that readers accessed both the figurative and the literal meanings of the metaphor simultaneously. In addition, the short timescale and the fact that subjects were engaged in a task unrelated to figurative meaning suggests that the figurative meaning was accessed automatically alongside the literal meaning. A similar study by Glucksberg and Keysar (1982), in which subjects indicated whether presented stimuli were literal or metaphoric also provided strong evidence that the literal and figurative meanings of metaphors are activated automatically and simultaneously.

The results of Glucksberg, Gildea, and Bookin (1982) and Glucksberg and Keysar (1982) made it clear the figurative meaning was not only accessed at the same time as the literal meaning, but also that the figurative meaning was accessed automatically, and not through the conscious, effortful process posited by SPM theory. Glucksberg eventually incorporated the idea that figurative meanings of metaphors are accessed automatically into a later model of metaphor processing, the Categorization model (Glucksberg, 2003; Glucksberg, 2008).

Glucksberg's (2003) Categorization model introduced several ideas about the structure of metaphors and the way that structure relates to metaphor comprehension. According to the Categorization model, the vehicle word of a metaphor holds "dual reference" (Glucksbery, 2003; 2008), which means it retains its standard literal definition but can also can act as a prototypical member of a category. In order for the metaphor to be properly understood, the appropriate figurative meaning of the metaphor vehicle word (inferred based on the category it represents) must be attributed onto the target. For instance, when reading the metaphor *love is a flower*, readers simultaneously access the literal meaning of flower as well as the relevant category (e.g., beautiful things) for which flower is acting as a prototype. They then attribute the relevant figurative characteristics of the vehicle word to the target word *love*, resulting in interpreting the metaphor *love is a flower* to mean, roughly, that love is beautiful. According to Glucksberg's Categorization model, both familiar and unfamiliar metaphors are processed in this way, although the categorical meaning of the target will be determined and processed more quickly and easily for familiar than unfamiliar metaphors.

The most recent major model of metaphor processing, Bowdle and Gentner's (2005) Career of Metaphor model, describes two different processing pathways for metaphors depending on their familiarity. According to the Career of Metaphor model, comprehension of familiar metaphors is an automatic process that is similar to the process described by the Categorization model, whereas comprehension of unfamiliar metaphors involves a controlled, effortful process of feature-mapping between the target and vehicle domains of the metaphor in order to determine the relevant characteristic that links the two domains (Bowdle and Gentner, 2005).

Several studies have yielded results supporting the Career of Metaphor model. Coney and Lange (2006) provided evidence that unfamiliar metaphors, metaphors which the reader has not previously encountered, may be processed differently than familiar metaphors. Coney and Lange used a vocal reaction time paradigm in which subjects read a series of unfamiliar metaphors, each of which was followed by the presentation of a prime word that was semantically associated with the metaphor they had just read. The time delay between presentation of the metaphor and presentation of the prime word varied between 375 ms, 750 ms, or 1500 ms across trials. Subjects were asked to name each word as quickly as possible, and faster-than-baseline naming times were taken as evidence of semantic priming from the metaphor. Coney and Lange found priming effects when the delay between metaphor presentation and presentation of the prime word was 1500 ms (sufficient time for some conscious processing), but not when it was 750 ms or 375 ms (sufficient time only for automatic processing). This pattern of results indicated that unfamiliar metaphors were not processed automatically, at least in the absence of supporting context. Rather, it seems that conscious attention was required to comprehend unfamiliar metaphors. This finding supports the Career of Metaphor model (Bowdle & Gentner, 2005) and seems to contradict Glucksberg's (2003; 2008) Categorization model, which does not differentiate between familiar and unfamiliar metaphors in terms of how they are processed.

Other studies have also supported the Career of Metaphor model's predictions regarding processing of familiar and unfamiliar metaphors. Lai, Curran, and Menn (2009) found larger amplitudes of the N400, an ERP component related to surprising or novel stimuli, when participants read unfamiliar metaphors than when they read familiar metaphors. Furthermore, ERPs of familiar metaphors closely resembled those of literal phrases, whereas ERPs of unfamiliar metaphors did not resemble those observed when reading literal phrases or familiar metaphors, even at later time windows. The researchers interpreted the results as being compatible with a view of unfamiliar metaphors being processed differently than familiar metaphors, with unfamiliar metaphors being cognitively taxing to comprehend. A recent eye-tracking study by Campbell (2014), which found longer gaze durations, more fixations, and more regressions in unfamiliar than familiar metaphors, also supports the Career of Metaphor model.

Clearly, several studies support the Career of Metaphor model's (Bowdle & Gentner, 2005) view that comprehension of unfamiliar metaphors is a cognitively taxing, effortful process while comprehension of familiar metaphors is automatic. However, a limitation of these studies is that they do not completely rule out a potential alternative explanation, which is that comprehension of both familiar and unfamiliar metaphors require effortful processing, but that this processing simply occurs much more quickly in familiar metaphors than unfamiliar metaphors. For instance, longer reading times for unfamiliar metaphors (Coney & Lange, 2006) or more fixations on unfamiliar metaphors (Campbell, 2014) might support a model that specifies separate comprehension processes for familiar and unfamiliar metaphors, but this result could also be taken as support for view in which comprehension of familiar and unfamiliar metaphors shares the same underlying process, but the process proceeds more quickly in familiar metaphors. Observed differences in ERPs for familiar and unfamiliar metaphors (Lai, Curran, & Menn, 2009) are also not incompatible with the possibility that the cognitive processes underlying familiar and unfamiliar metaphor comprehension are fundamentally similar, but differ radically in timescale.

To definitively conclude that separate processes are used to comprehend familiar and unfamiliar metaphors, the possibility of a shared underlying comprehension process operating at different timescales must be ruled out. To do so, one must be able to measure the effects of effortful processing separately from automatic processing. If effortful processes were shown to affect comprehension of unfamiliar, but not familiar metaphors, this would provide support for the idea that the processes underlying comprehension of familiar and unfamiliar metaphors are dissociable, and would thus provide further support for the Career of Metaphor model (Bowdle & Gentner, 2005). A possible example of a process that fits this description is resource depletion (Baumeister, Bratslavsky, Muraven, & Tice, 1998).

B. **Resource Depletion**

Resource depletion theory (Muraven & Baumeister, 2000; Baumeister, Vohs, & Tice, 2007) views self-control, also called cognitive control, as a limited cognitive resource. Research demonstrates (e.g., Baumeister, Bratslavsky, Muraven & Tice, 1998; Smit, Eling & Coenen, 2004) that exerting self-control on an initial task leads to impaired performance on a subsequent task that requires self-control. The nature and degree of the performance decrement varies depending on the domain being observed.

The basic design of resource depletion studies is simple--subjects are assigned to complete a depletion task or non-depletion task (common tasks are described below). The depletion and the non-depletion groups then complete a second task. Performance on the second task is the measure of interest. In the depletion condition, the first task (depletion task) is difficult and requires the exertion of cognitive control whereas in the non-depletion condition the first task (non-depletion task) requires substantially less cognitive control. Performance on the second task is expected to be reduced for the depletion group relative to the non-depletion group, but only if the second task requires controlled processing.

A common depletion task involves exerting cognitive control by controlling or suppressing emotions. In one such study by Baumeister, Bratslavsky, Muraven and Tice (1998), participants in the depletion condition watched an emotionally charged video and were instructed to suppress their emotional responses to the video. Participants in the non-depletion (control) condition were instructed to simply watch the video with no mention of suppressing emotions. Both groups completed an anagram-solving task after watching the video. Baumeister et al. (1998) found that participants in the non-depletion condition performed significantly better on the anagram-solving task compared to participants in the depletion condition, an effect which they attributed to the participants' differential exertion of cognitive control while watching the video.

From an initial focus on emotional self-control, research on resource depletion has expanded to examine self-control in a variety of social and cognitive contexts. In a study by Vohs and Heatherton (2000), dieters watched a neutral video in one of two conditions. In the high-temptation (depletion) condition, a bowl of M&Ms was present on a table right next to them as they watched the video. In the low-temptation (control) condition, the bowl of M&Ms was placed further away. Participants were not explicitly instructed to not eat the M&Ms, as the study relied on the fact that the participants, who were currently dieting, were internally motivated avoid eating sugary and unhealthy foods. Participants in the high-temptation condition persisted less on a subsequent unsolvable embedded figures task compared to their counterparts in the low-temptation condition. These results indicated that exercising cognitive control in the form of resisting temptation produced a depletion effect. Resource-depletion effects have also been observed in a variety of other social contexts, such as interaction with other-race individuals (Richeson &Shelton, 2003), acting counter to gender norms (Vohs, Baumeister, & Ciarocco, 2005), and consumer decision-making (Vohs et al., 2008).

Resource depletion has also been observed to differentially affect performance of tasks that require active self-regulation compared to those which do not. In three related experiments, Schmeichel, Vohs, and Baumeister (2003) asked participants to view a short, silent video of a woman being interviewed for a job in which a list of unrelated words was shown one-at-a-time at the bottom of the video. Participants in the depletion condition were instructed to make an effort to not look at the words. Participants in the control condition were given no instruction regarding the words in the video. Schmeichel et al. measured performance on either simple information processing tasks (e.g., rote memorization of nonsense syllables) or tasks requiring complex logical reasoning (e.g., analytical GRE problems). Schmeichel et al. found that depletion induced poorer cognitive performance on the complex reasoning tasks relative to the non-depletion group, but not on simple information retrieval tasks, suggesting that complex cognition requires active self-regulation and can thus be affected by the previous exertion of cognitive control.

Another corollary of the limited resource theory of cognitive control is the possibility of training cognitive control analogously to the way a muscle is trained. Oaten and Cheng (2006a, 2006b; 2007) have demonstrated that a diverse range of training techniques involving cognitive control, from regular physical exercises to rigorous programs of study or personal financial monitoring over a period of a few months, can increase the capacity to exert cognitive control on unrelated tasks. This supports the idea of a common resource for cognitive control that can be depleted by a wide range of diverse tasks and can be improved by training on a diverse range of self-control tasks. In a meta-analysis of 83 published studies, Hagger et al. (2010) found significant effects of depletion on self-control task performance across different kinds of depletion tasks and subsequent dependent tasks (the second task). These findings highlight the wide range of contexts in which resource depletion has been observed, and provide evidence for cognitive control as a limited resource that is used across a variety of contexts.

C. **Current Study**

The research described above demonstrates that resource depletion can impair cognitive performance on tasks that require controlled processing. The study proposed here explores whether resource depletion can affect the processing of metaphors. Given that resource depletion effects appear to be domain-general, having been observed in cognitive (e.g., Schmeichel, Vohs, & Baumeister, 2003), affective (e.g., Baumeister et al., 1998), and social tasks (e.g., Richeson & Sheldon, 2003), there appears to be no fundamental reason that depletion effects could not be observed in the domain of metaphor processing. The primary research question is, does metaphor processing involve the exertion of cognitive control? If the answer is yes, depletion of a cognitive control resource should impair metaphor processing.

Given that major models of metaphor processing specify that familiar metaphors are processed automatically (Glucskberg, 2003; Bowdle & Gentner, 2005), it is unlikely that comprehending familiar metaphors requires the exertion of cognitive control. And given the Career of Metaphor (Bowdle & Gentner, 2007) model's view of unfamiliar metaphor comprehension as an effortful process that requires active attention, it seems probable that processing of unfamiliar metaphors requires the exertion of cognitive control. The analytical feature-mapping process described in the Career of Metaphor model appears to be similar to the higher-order analytical tasks used by Schmeichel et al. (2003). In Schmeichel et al., performance on higher-order analytical tasks was impaired by resource depletion whereas performance on rote recall tasks was not affected by resource depletion. In sum, it appears plausible that depletion of a cognitive control resource would affect processing of unfamiliar metaphors, which require cognitive control, but not familiar metaphors, which are processed in a more automatic manner.

If resource depletion does affect or impair the processing of unfamiliar metaphors, a further question is, how will this impairment be observed? Based on resource depletion effects in other domains, resource-depleted readers should have greater difficulty in arriving at the meaning of an unfamiliar metaphor. This difficulty could be observed in comprehension times for unfamiliar metaphors. If resource-depleted participants took longer than non-depleted participants to comprehend unfamiliar metaphors, but not familiar metaphors, this would support different processes comprehension for familiar and unfamiliar metaphors.

The experiments described below represent an exploratory study examining the effects of resource depletion on the processing and comprehension of metaphors. To investigate how resource depletion affects processing of metaphors, I conducted two experiments in which participants were assigned to either the control (non-depletion) condition or the resource depletion condition. Across both experiments, the depletion manipulation is a Stroop color-word naming task (MacLeod, 1991) in which participants are presented with words that are color names and are instructed to name the font color that each word is presented in. Completing a Stroop color-word task requires the exertion of cognitive control and self-regulation to varying degrees, depending on the percentage of congruent and incongruent trials presented (Hagger et al., 2010). Incongruent trials present the name of a color in a colored typeface that does not match the named color (e.g., the word “blue” presented in red font). As such, incongruent trials require that readers suppress a well-learned prepotent response, saying the word name, in order to perform the task of saying the color of the ink (Hagger et al., 2010). Conversely, congruent trials present color names in a font color that matches the name (e.g., the word “blue” presented in blue font). Congruent trials do not require the exertion of cognitive control because no response needs to be inhibited.

All participants performed a 120-trial Stroop color-word task. In the control condition, participants named the word, whereas in the resource depletion condition, participants named the color that the word was written in. As such, the amount of cognitive control exerted by participants in the control condition should be minimal whereas the amount of cognitive control exerted by participants in the depletion condition should be substantial. In both the conditions, 70% of the trials were incongruent and 30% congruent. The Stroop task has been used as a depleting task in several depletion studies (e.g., Webb & Sheeran, 2003; Bray et al., 2008). These studies used Stroop tasks with between 60 and 80 percent incongruent trials, as this ratio induces depletion both by exposing participants to many incongruent trials and by not allowing participants to settle into a response pattern for either congruent or incongruent trials (Hagger et al., 2010). In a meta-analysis, Hagger et al. (2010) found that the Stroop task produces highly consistent, moderately-sized depletion effects. In addition, because participants must continuously respond, an advantage of the Stroop-color word task is that it is less affected by variations in degree of subject motivation and adherence to instruction as compared to affective depletion tasks such as watching videos (Hagger et al., 2010).

In Experiment 1, I investigated the effects of resource depletion on the amount of time needed to comprehend metaphoric and literal sentences. After the depletion or non-depletion manipulation, participants were presented with a series of high-familiarity metaphors (e.g., *the mind is a sponge*), mid-familiarity metaphors (e.g., *sleep is an ocean*), and low-familiarity metaphors (e.g., *a fisherman is a spider*), as well as literal phrases that structurally match the metaphor phrases (e.g., *a building is a skyscraper*). Participants were instructed to press a response button as soon as they understand each phrase, and response times were recorded. Given that low-familiarity metaphors take longer to comprehend than high-familiarity metaphors, response times on this task are known to reflect processing differences between high-and low-familiarity metaphors. The key question is, is the difference in response times for participants in the depletion and non-depletion conditions similar for high-, mid, and low-familiarity metaphors? The following hypotheses were tested in Experiment 1:

1. A main effect of Stimulus Group: response times will be longest for low-familiarity metaphors, followed by mid- and high-familiarity metaphors, with literal sentences having the fastest response times. Response times for familiar metaphors might be slower than response times for literal statements because, despite being normed to be familiar, every subject might not be highly familiar with every familiar metaphor whereas the meaning of literal statements is obvious.
2. A main effect of depletion condition: response times will be longer in the resource depletion condition than in the control condition.
3. A Depletion condition X Stimulus Group interaction. Specifically, this prediction relates to the three groups of metaphor familiarity. Differences in response times between the control and depletion conditions will be largest for low-familiarity metaphors, next largest

for mid-familiarity metaphors, smaller for high-familiarity metaphors, and smallest for literal statements.

Experiment 2 investigated the effects of resource depletion on metaphor processing using a task that reflects processing differences between metaphors and literal phrases, but not between different degrees of familiarity within metaphors phrases. After the initial depletion or non-depletion manipulation, participants were presented with a series of high-, mid-, and low-familiarity metaphors, as well as literal phrases that structurally match the metaphor phrases (the same stimuli as Experiment 1). For each sentence, participants pressed a response button to indicate whether the meaning of the presented sentence was concrete or abstract as quickly and accurately as possible. The rationale for this decision task is that participants do not need to fully comprehend the metaphors to make their response – they only need to decide whether or not the sentence they are reading has a concrete meaning (this category corresponds to literal sentences) or an abstract meaning (this category corresponds to metaphor sentences regardless of familiarity) and categorize it accordingly. As such, this task should not be very sensitive to differences in processing between metaphors of different levels of familiarity. This conclusion is also supported by pilot data, which I discuss after presenting the primary hypotheses. Once again, the key question for Experiment 2 is, is the difference in response times for participants in the depletion and non-depletion conditions similar for high-, mid-, and low-familiarity metaphors? The following hypotheses were tested in Experiment 2.

1. A main effect of Stimulus Group: response times will be slightly longer for metaphor stimuli than for literal stimuli, but there will be no difference in response time among the different familiarity metaphor conditions.
2. A main effect of depletion condition: response times will be longer in the resource depletion condition than in the control condition. The magnitude of the depletion effect is

hypothesized to be smaller than in Experiment 1, as the response task in Experiment 2 requires less exertion of cognitive control.

3. No Stimulus Group X Depletion condition interaction: there will be no difference in the size of the resource depletion effect between low-, mid-, and high-familiarity metaphors as well as literal statements.

In Experiment 1, the response task is expected to be sensitive to processing differences between metaphors of varying familiarity. In Experiment 2, the response task is not expected to be sensitive to processing differences based on metaphor familiarity. If the depletion task differentially affects processing of high- and low- familiarity metaphors in Experiment 1 but not in Experiment 2, this result would support the conclusion that the processes that underlie comprehension of familiar and unfamiliar metaphors are dissociable (different). This would provide empirical support for the Career of Metaphor model, which specifies different processes for familiar and unfamiliar metaphors.

In both experiments, the metaphors were drawn from a large bank of metaphors normed by Katz et al. (1988) on metrics of familiarity, number of alternative interpretations, and other characteristics. A recent replication (Campbell & Raney, 2015) of Katz et al.'s norming study yielded similar ratings on familiarity and the other metrics recorded in Katz et al., which confirms the validity of the norms for the UIC student population.

D. **Pilot Studies**

The methodology is based on three pilot studies, of which pilot studies 1 and 3 closely match Experiments 1 and 2, respectively. These pilot studies were conducted to test the effectiveness of using resource depletion to assess differences in metaphor processing based on familiarity. The pilot studies used a different depleting task from the one used in the current studies (this depleting task is detailed in the Method section of each pilot study, presented in

Appendix C). In all three pilot studies, the stimulus set consisted of 60 short phrases, of which 30 were literal phrases (e.g., *a skyscraper is a building*), 15 were high-familiarity metaphors (e.g., *the mind is a sponge*), and 15 were low-familiarity metaphors (e.g., *a fisherman is a spider*). High-familiarity was defined as being in the top quartile of Katz' normed familiarity ratings, and low-familiarity was defined as being in the bottom quartile of Katz' normed familiarity ratings.

In pilot study 1, participants read the sentences and pressed a response key as soon as they understood the sentence (the same instruction as proposed for Experiment 1). In pilot study 2, participants used response keys to indicate, as quickly and accurately as possible, whether each presented sentence was literal or figurative. In pilot study 3, participants indicated whether the meaning of each sentence was abstract or concrete, which does not require subjects to fully comprehend the metaphors (similar to the task proposed for Experiment 2). A video task (described in detail in Appendix C) was used to induce depletion. In brief, the video showed a woman interviewing for a job with a list of unrelated words displayed in the bottom of the video. The video was silent. Participants in the depletion conditions were instructed to ignore the words (not look at them). Participants in the non-depletion conditions were given no instructions about the words except that unrelated words would appear in the video. Thus, for each pilot the independent variables were depletion condition (depletion, no depletion) and stimulus group (literal statements, low familiarity metaphors, high familiarity metaphors). Note that medium-familiarity metaphors were not included in the pilot studies. The only difference between the pilot studies was the decision made in the second task. To review, in pilot 1, participants responded when they knew the meaning of the statement. In pilot 2, participants made a figurative/literal judgment. In pilot 3, participants made a concrete/abstract judgment. Response time was measured for each pilot study. A detailed description of the methodology for all three pilot studies is presented in Appendix C. The results of pilot studies 1, 2, and 3 are summarized below.

1. **Pilot Study 1**

A mixed-model ANOVA was performed (n = 16) with response times as the dependent measure, depletion condition (depletion, non-depletion) as the between-subjects measure, and stimulus group (low familiarity metaphors, high familiarity metaphors, literal statements) as the within-subjects measure. Results indicated a main effect of stimulus group ($p < .001$). Literal statements ($M = 1491$ ms) were understood more quickly than high-familiarity metaphors ($M = 2866$ ms), which in turn were understood more quickly than low-familiarity metaphors ($M = 4677$ ms). Results also indicated a main effect of depletion condition ($p = .04$): resource depleted participants had longer response times ($M = 3556$ ms) than participants in the control condition ($M = 2466$ ms). When running the analysis with only the low-familiarity metaphor and high-familiarity metaphor stimulus groups (i.e., no literal statements, as these were not relevant to the hypothesized interaction), the Stimulus Group X Depletion Condition interaction was not significant ($p = .276$). While not significant, the expected direction of the interaction was observed, as the magnitude of the depletion effect (defined as the difference between the depletion and non-depletion conditions) was 1917 ms for low-familiarity metaphors and 1096 ms for high-familiarity metaphors. The observed power for this interaction effect was very low (0.19).

The results of pilot study 1 provide evidence that resource depletion affects the processing of metaphors. Given the effectiveness of this methodology, Experiment 1 closely matches this design, with several additions. Most notably, a “mid-familiarity metaphor” stimulus condition was added and a much larger sample of participants was recruited (described in detail below). This will increase the statistical power for detecting a Depletion X Stimulus Group interaction.

2. **Pilot Study 2**

A mixed-model ANOVA was performed (N = 40) with response times as the dependent measure, depletion condition as the between-subjects measure (depletion, non-depletion), and stimulus group (low familiarity metaphors, high familiarity metaphors, literal statements) as the within-subjects measure. Results indicated a main effect of stimulus condition ($p < .001$). Literal statements were responded to more quickly than either metaphor type, but there was no difference in response times between high-familiarity (M = 1365 ms) and low-familiarity metaphors (M = 1402 ms). Results indicated no main effect of depletion condition ($p = .24$) and no Stimulus Group X Condition interaction ($p = .63$).

3. **Pilot Study 3**

Because no depletion effect (main effect of condition) was observed in pilot study 2, whereas a depletion effect was observed when using the same video task as a depleting task in pilot study 1, I reasoned that the metaphor/literal categorization task did not require participants to exert cognitive control to a significant degree. Thus, I chose a new decision task for pilot study 3 that fit the criteria of requiring the exertion of cognitive control while also not imposing different processing demands for familiar and unfamiliar metaphors. For pilot study 3, participants were asked to decide as quickly as possible whether the meaning of each sentence was concrete or abstract. As such, pilot study 3 was identical to pilot study 2 and pilot study 1 with the exception of the response decision. A mixed-model ANOVA was performed (n = 20) with response times as the dependent measure, Depletion Condition as the between-subjects measure, and Stimulus Group as the within-subjects measure. Results indicated a main effect of Stimulus Group ($p < .001$). Literal statements were responded to more quickly than either metaphor type, but there was no significant difference in response times between high-familiarity (M = 1321 ms) and low-familiarity (M = 1285 ms) metaphors. Results indicated a main effect of Depletion Condition (*p* < .05), and no Stimulus Group X Condition interaction (*p* = .30).

These results indicate that the abstract/concrete decision task used in pilot study 3 met the desired criteria. I observed a depletion effect (main effect of condition), indicating that the task required the exertion of cognitive control. I also observed no difference in response times or size of the depletion effect between familiar and unfamiliar metaphors, indicating that the abstract/concrete task did not impose different processing demands for familiar and unfamiliar metaphors. As such, this response task was used in Experiment 2.

The methodology for Experiments 1 and 2 is identical except for the decision made; therefore, a combined methods section is presented below.

II. METHOD

A. **Participants**

A total of 120 participants were recruited for Experiments 1 and 2 (60 for each experiment). Participants were undergraduate students in the University of Illinois at Chicago (UIC) psychology subject pool who volunteered for the study in exchange for course credit. Participants were required to be proficient English speakers, which was defined as attending an English-speaking school for at least 10 years. To assess language backgrounds and vocabulary knowledge among participants, all participants were administered a language history questionnaire and vocabulary quiz (described below).

B. **Materials**

A total of 90 sentence stimuli were created. All sentences were short phrases of the format "A(n) X is a Y," which are grammatically acceptable as standalone sentences. Of these, 15 were low-familiarity metaphors, 15 were mid-familiarity metaphors, 15 were high-familiarity metaphors, and 45 were literal comparison sentences constructed to match the structure of a metaphor. An example of a matching literal phrase is "a flamingo is a bird," wherein the pseudo-target word "flamingo" corresponds to the target word in a metaphor and the pseudo-vehicle word "bird" corresponds to the vehicle word in a metaphor. To avoid potential confounds of word length and word frequency, sentences were created and selected such that the average word length and word frequency of the target/pseudo-target word in each stimulus group was roughly equal, and the average word length and word frequency of the vehicle/pseudo-vehicle word in each stimulus group was roughly equal. Sentences were presented in a pseudorandomized order, manipulated such that no one type of sentence (high-familiarity metaphor, low-familiarity metaphor, mid-familiarity metaphor, or literal) appears more than three times consecutively. A table of sample stimuli is presented in Table I, Appendix A. Sentences were displayed on a 15-

inch monitor in 12pt black font on a white background, with participants maintaining a viewing distance of approximately 20 inches.

Katz's (1988) database contains 256 metaphors normed on a set of ten criteria including familiarity. High-familiarity metaphors were selected from among metaphors in the top 25% of familiarity ratings. Low-familiarity metaphors were selected from metaphors in the bottom 25% of familiarity ratings. Mid-familiarity metaphor were chosen from those falling within +/- 12.5% of the median rating of familiarity. Previous research on metaphor comprehension (e.g. Lai et al., 2009) has usually used two metaphor familiarity groups, familiar and unfamiliar, which would correspond to the high- and low-familiarity groups in this study. A potential advantage of including a third familiarity group in between the polar extremes of familiarity is gaining additional precision in describing how comprehension processes change as familiarity changes from. For instance, by including mid-familiarity metaphors, I may be able to observe whether comprehension difficulty increases linearly or non-linearly as familiarity decreases.

To avoid potential confounds of word length and word frequency, sentences were created and selected such that the average word length and word frequency of the target/pseudo-target word in each stimulus group was roughly equal, and the average word length and word frequency of the vehicle/pseudo-vehicle word in each stimulus group was roughly equal. Word frequencies were determined using the Corpus of Contemporary American English (COCA, http://corpus.byu.edu/coca/).

In both experiments, the depletion manipulation used a Stroop color-word task with 120 trials, consistent with the duration of previous research using the Stroop task as a depletion task (e.g., Webb & Sheeran, 2003). On each trial, participants saw a word corresponding to a color (e.g., "blue") presented in a certain font color. On congruent trials, the font color matched the word (e.g., the word "blue" presented in blue font). On incongruent trials, the font color did not match the word (e.g., the word "blue" presented in red font). Participants in both conditions performed a version of the task consisting of 70% incongruent and 30% congruent trials. Words were presented in Arial 20pt font (green, red, and blue) on a 15-inch display, and participants maintained a viewing distance of about 20 inches. Participants in the control condition were instructed to respond by indicating the word which was presented, whereas participants in the depletion condition were instructed to respond by indicating the font color that the word is written in. Participants indicated their responses by pressing one of three color-coded buttons, corresponding to the three colors used in the experiment, on a Cedrus RB series response pad.

The Metaphor Familiarity Rating task in this experiment presented participants with each of the 45 metaphor sentences encountered earlier in the study. For each metaphor, the participant is instructed to rate how familiar the metaphor is to them on a 1-7 scale (1 = very unfamiliar, 7 = very familiar). In addition, the participant is instructed to write their best interpretation of the meaning of each metaphor in 1-2 sentences. The Metaphor Familiarity Rating task was presented on a computer via an online Qualtrics survey, and results were collected and stored online in the Qualtrics system.

The UIC English Vocabulary quiz was used in this experiment to measure English word knowledge. Developed by Raney, the test consists of 30 multiple-choice items for which the participant is instructed to choose the best answer among five alternatives. This vocabulary test has been used in prior studies (Minkoff & Raney, 2000; Therriault & Raney, 2007), and has been found to be correlated with comprehension skill (r = 0.40 to 0.52). The Vocabulary Quiz is presented in Appendix B. Vocabulary scores were compared across groups to ensure that that participants in the depletion and non-depletion conditions had similar vocabulary scores (averages within 1 point of each other).

The Language History Questionnaire consists of several questions asking participants to list which languages they know, which language they learned first, and rate their proficiency in speaking, comprehending, and reading each of the languages they know. The Language History

Questionnaire is presented in Appendix B. Responses to the language history questionnaire were used to verify that participants in the depletion and non-depletion conditions had similar language backgrounds (e.g. ensuring that depletion and control conditions had a similar number of native English speakers).

C. **Procedure**

Participants performed the experiment in a quiet room in groups of one to three. The experiment included two main tasks: the Stroop depletion manipulation and the metaphor task. Participants began by signing an informed consent form. They then completed the UIC English Vocabulary Quiz and the Language History Questionnaire, which were presented via an online Qualtrics survey. After participants complete these tasks, the experimenter gave a brief verbal explanation of the Stroop color-word task.

The experimenter instructed participants in the depletion condition by saying: “In this experiment, you will be presented with a number of words. The word will be presented in different font colors. For each word, you will need to press the key on the button box that corresponds to the color of the font, not the word itself. For example, if you see the word ‘blue’ written in green font, you should respond ‘green’.” For participants in the non-depletion condition, the experimenter said: “In this experiment, you will be presented with a number of words. The words will be presented in different font colors. For each word, you will need to press the key on the button box that corresponds to the word itself, not the color of the font. For example, if you see the word “blue” written in green font, you should respond “blue”.” The experimenter also answered any questions from the participant before starting the Stroop task. Participants generally finished the Stroop task at around the same time (within 1 minute of each other) in a given session. Participants who finished early were instructed to sit quietly and wait until all participants had finished. After participants finished the Stroop task, the experimenter gave verbal instructions for the metaphor and literal sentence comprehension task.

In Experiment 1, participants indicated the moment at which they understood the presented sentence or arrived at a reasonable interpretation of its meaning by pressing a single response key on a response pad. I refer to this as the comprehension task. Participants were instructed as follows: "In this task you will read a number of short sentences for comprehension. For each sentence, press the response button as soon as you have understood the meaning of the sentence". In Experiment 2, participants indicated whether the sentence is "concrete" or "abstract" by pressing one of two labeled response keys on the response pad. I refer to this as the concrete/abstract task. Participants were instructed as follows: "In this task you will read a number of short sentences. For each sentence, you will need to decide, as quickly as possible, whether the meaning of that sentence is concrete or abstract. The meaning of a sentence is concrete if it is referring to objects or relationships that are physically real. For instance, the meaning of the sentence *a hammer is a tool* is concrete. The meaning of a sentence is abstract if it refers to ideas or relationships that are not physically real. For instance, the meaning of the sentence sentence *love is a flower* is abstract. Respond "concrete" or "abstract" based on the meaning of the sentence as a whole, not on the individual words in the sentence. If the meaning is concrete, press the response button labeled "concrete". If the meaning is abstract, press the response button labeled "abstract"."

The Superlab program (Cedrus, 2010) was used to control all aspects of the experiment, such as timing of stimuli presentation and recording decision times. After each trial, participants were able to advance to the next trial by pressing the spacebar, whereupon a fixation cross was presented in the center of the screen for 500 milliseconds, followed by presentation of the next phrase.

Five practice trials were presented at the beginning of the task to allow participants to become familiar with the task. After completing the response task, participants were instructed to complete the Metaphor Familiarity Rating task on the computer via Qualtrics online survey, and were instructed to try to base their assessments of each metaphor's familiarity on how familiar it was to them prior to participating in this study. This instruction was intended to minimize potential carryover effects due to the participants having already encountered each metaphor earlier in the study. After completing the Metaphor Familiarity Rating task, participants were given a written debriefing form describing the study's goals. In both Experiment 1 and Experiment 2, the typical duration of each experimental session was 45-55 minutes and did not exceed one hour.

III. RESULTS AND DISCUSSION

A. Experiment 1 - Results

Results for Experiment 1 showed a great deal of variation both within any given participant's response times and between participants. In addition, because participants were not subject to a time limit in pressing the response key to indicate they had understood each sentence, the presence of trials with very long response times biased averages towards these outliers. To mitigate the effect of outlier response times, I excluded trials with response time of over 20 seconds from all analyses. A small number of very short response times were also observed, likely indicating a stray press of the response key. To account for these outliers, I also excluded all trials under 400 milliseconds, as response times below this threshold could not correspond to deliberate presses in response to the presented sentence.

To account for both subject and item variance, I conducted ANOVAs of decision times based on both subject means (F1) and item means (F2). In addition, I conducted a linear mixed-effects model (LME) analysis to combine subject and item variance to confirm the results of the ANOVAs. When describing differences between conditions, means are presented based on the by-subjects analyses only because the means by subjects and means by items showed the same patterns and absolute differences between the subject means and item means were small.

For the by-subject analysis, I conducted a 2 (Depletion Condition: Control, Depletion) x 4 (Stimulus Group: Literal, High-Familiarity Metaphor, Mid-Familiarity Metaphor, Low-Familiarity Metaphor) mixed-design ANOVA on mean per-subject comprehension times (in ms). For the by-item analysis, the same analysis was conducted with Depletion Condition as the within-item factor and Stimulus Group as the between-item factor.

As described below, the analysis revealed a significant main effect of Stimulus Group, a main effect of Depletion Condition, and a Stimulus Group X Depletion Condition interaction. The main effect of Depletion Condition was significant in both the by subject analysis, *F1*(1, 78) = 4.03, MS_e = 4234313, *p* = .048, ω^2 = .05, and in the by-item analysis, *F2*(1, 86) = 91.15, MS_e = 83400, *p* < .001, ω^2 = .52. Average comprehension time for the Control condition (*M* = 2425 ms, *SD* = 1423) was 461 ms faster than the Depletion condition (*M* = 2986 ms, *SD* = 1601). The main effect of Stimulus Group was also significant in both the by-subject analysis, *F1*(3, 234) = 111.42, MS_e = 690582, *p* < .001, ω^2 = .59, and the by-item analysis, *F2*(3, 86) = 187.42, MS_e = 252630, *p* < .001, ω^2 = .87. Pairwise comparisons revealed that all differences between pairs of stimulus groups were significant. When ordered from fastest to slowest comprehension times, comprehension times for Literal sentences (*M* = 1576 ms, *SD* = 601) were 819 ms faster than for High-Familiarity metaphors (*M* = 2395 ms, *SD* = 935), *p* < .001. Comprehension times for High-Familiarity metaphors were 833 ms faster than Mid-Familiarity metaphors (*M* = 3228 ms, SD = 1518), *p* < .001. Comprehension times for Mid-Familiarity metaphors were 595 ms faster than Low-Familiarity metaphors (*M* = 3823 ms, SD = 1728), *p* < .001. Finally, the Stimulus Group X Depletion Condition interaction was also significant in both the by-subject analysis, *F1*(3, 234) = 3.18, MS_e = 690582, *p* = .025, ω^2 = .04, and the by-item analysis, *F2*(3, 86) = 14.36, MS_e = 83400, *p* < .001, ω^2 = .33. In the by-subject analysis, observed power was 1.00 for the main effect of Stimulus Group, .51 for the main effect of Depletion Condition, and .73 for the interaction. In the by-item analysis, observed power was 1.00 for both main effects and for the interaction.

To probe the Stimulus Group X Depletion Condition interaction, I tested the simple effects of Depletion Condition at each level of Stimulus Group, based on the data from the by-subject analysis. When responding to Low-Familiarity Metaphors, participants in the Control condition (*M* = 3445 ms, *SD* = 1736) were 756 ms faster on average than participants in the Depletion condition (*M* = 4201 ms, *SD* = 1656), *F*(1, 78) = 3.977, MS_e = 2877068, *p* = .049, ω^2 = .05. When responding to Mid-Familiarity metaphors, participants in the Control condition (*M* = 2861 ms, *SD* = 1350) were 734 ms faster on average than participants in the Depletion condition, (*M* =

3595 ms, *SD* = 1603), *F*(1, 78) = 4.91, MS_e = 2195563, *p* = .03, ω^2 = .06. When responding to High-Familiarity metaphors, comprehension times were not significantly different between participants in the Control Condition (*M* = 2276 ms, *SD* = 1004) and participants in the Depletion condition, (*M* = 2514 ms, *SD* = 857), *F*(1, 78) = 1.31, MS_e = 870881, *p* = .26, ω^2 = .02. Finally, when responding to Literal sentences, comprehension times were not significantly different between participants in the Control condition (*M* = 1517 ms, *SD* = 564) and participants in the Depletion condition (*M* = 1634 ms, *SD* = 638), *F*(1, 78) = .76, MS_e = 362549, *p* = .39, ω^2 = .01. To summarize the comparisons, there was a significant depletion effect for Low- and Mid-familiarity metaphors, but not for High-familiarity metaphors or literal statements.

Because the interaction I hypothesized referred specifically to metaphors of different familiarity levels and not literal statements, I performed the same analyses with only the three metaphor familiarity groups (literal stimuli were excluded). This analysis again revealed a Stimulus Group X Depletion Condition interaction, which was significant in both the by-subject analysis, *F1*(2, 156) = 3.31, MS_e = 517635, *p* = .04, ω^2 = .04, and the by-item analysis, *F2*(2, 42) = 4.18, MS_e = 147052, *p* = .02, ω^2 = .17. This demonstrates the predicted interaction is not dependent on the inclusion of the literal condition. The analyses also revealed a main effect of Stimulus Group, which was significant in both the by-subject analysis, *F1*(2, 156) = 79.55, MS_e = 517635, *p* < .001, ω^2 = .50, and the by-item analysis, *F2*(2, 42) = 38.36, MS_e = 398555, *p* < .001, ω^2 = .65, and a main effect of Depletion Condition, which was significant in both the by-subject analysis, *F1*(1, 78) = 4.06, MS_e = 4908241, *p* = .047, ω^2 = .05, and the by-item analysis, *F2*(1, 42) = 50.44, MS_e = 147052, *p* < .001, ω^2 = .55. Follow-up analyses are not reported here as the pattern of results is the same as in the ANOVA analyses with all four groups of stimuli included.

Linear Mixed-Effect Model Analysis

I also conducted a linear mixed effect model (LME) analysis (trials nested within individual participants) on comprehension times as a function of Depletion Condition (Control, Depletion) and metaphor Stimulus Group (High-Familiarity, Mid-Familiarity, Low-Familiarity). The linear mixed effect model is useful because analyses can be set up to simultaneously include subject and item variability. If the ANOVAs reflect a spurious result due to unexpected subject or item variability, or if the LME model has increased statistical power, the LME model could produce a different pattern of results. To match the ANOVA analyses, I performed two LME analyes, one with all four stimulus groups included and one with only the three metaphor stimulus groups included.

As with the corresponding ANOVA model, the LME analysis with all four stimulus groups included revealed a main effect of Depletion Condition: comprehension times for the Control condition (M = 2192 ms, SD = 1933) were 341 ms faster than for the Depletion condition (M = 2533 ms, SD = 2324), $F(1, 2769) = 46.26$, $p < .001$. The LME analysis also revealed a main effect of Stimulus Group, $F(2, 1838) = 112.71$, $p < .001$. Pairwise comparisons, presented from fastest to slowest response times, revealed that response times for Literal sentences (M = 1580 ms, SD = 1149) were 821 ms faster than High-Familiarity metaphors (M = 2401 ms, SD = 1779), $p < .001$, High-Familiarity metaphors were 836 ms faster on average than for Mid-Familiarity metaphors (M = 3237 ms, SD = 2594), $p < .001$, and response times for Mid-Familiarity metaphors were 581 ms faster on average than for Low-Familiarity metaphors (M = 3818 ms, SD = 3013), p < .001. Importantly, the Depletion Condition X Stimulus Group interaction was significant, $F(3, 1244) = 8.90$, $p < .001$.

To follow up the Depletion Condition X Stimulus Group interaction, I tested the simple effect of Depletion Condition at each level of Stimulus Group. The simple effect of Depletion Condition was significant at every level of Stimulus Group. In Literal sentences, Depletion condition participants (M = 1638, SD = 1226) had longer comprehension times by 117 ms than Control condition participants (M = 1521, SD = 1064), $F(1, 2646) = 6.16$, $p = .01$. In High-Familiarity metaphors, Depletion condition participants (M = 2524 ms, SD = 1797) had longer response times by 244 ms than Control condition participants ($M = 2280$ ms, $SD = 1753$), $F(1, 1035) = 5.67$, $p = .02$. In Mid-Familiarity metaphors, Depletion condition participants ($M = 3609$ ms, $SD = 2891$) had longer response times by 742 ms on than Control condition participants ($M = 2867$ ms, $SD = 2201$), $F(1, 1054) = 19.84$, $p < .001$. Finally, in Low-Familiarity metaphors, Depletion condition participants ($M = 4184$ ms, $SD = 3185$) had longer response times by 732 ms than Control condition participants ($M = 3452$ ms, $SD = 2787$), $F(1, 978) = 17.43$, $p < .001$.

The LME analysis with only the three metaphor stimulus groups revealed a main effect of Depletion Condition: comprehension times for the Control condition ($M = 2865$ ms, $SD = 2333$) were 575 ms faster than for the Depletion condition ($M = 3440$ ms, $SD = 2777$), $F(1, 2661) = 42.47$, $p < .001$. The LME analysis also revealed a main effect of Stimulus Group, $F(2, 1838) = 112.71$, $p < .001$. Pairwise comparisons, presented from fastest to slowest response times, revealed that response times for High-Familiarity metaphors ($M = 2401$ ms, $SD = 1779$) were 836 ms faster on average than for Mid-Familiarity metaphors ($M = 3237$ ms, $SD = 2594$), $p < .001$, and response times for Mid-Familiarity metaphors were 581 ms faster on average than for Low-Familiarity metaphors ($M = 3818$ ms, $SD = 3013$), p < .001. Importantly, the Depletion Condition X Stimulus Group interaction was significant, $F(2, 1838) = 4.21$, $p = .02$.

To follow up the Depletion Condition X Stimulus Group interaction, I tested the simple effect of Depletion Condition at each level of Stimulus Group. The simple effect of Depletion Condition was significant at every level of Stimulus Group. In High-Familiarity metaphors, Depletion condition participants ($M = 2524$ ms, $SD = 1797$) had longer response times by 244 ms than Control condition participants ($M = 2280$ ms, $SD = 1753$), $F(1, 1035) = 5.67$, $p = .02$. In Mid-Familiarity metaphors, Depletion condition participants ($M = 3609$ ms, $SD = 2891$) had longer response times by 742 ms on than Control condition participants ($M = 2867$ ms, $SD = 2201$), $F(1, 1054) = 19.84$, $p < .001$. Finally, in Low-Familiarity metaphors, Depletion condition participants

(*M* = 4184 ms, *SD* = 3185) had longer response times by 732 ms than Control condition participants (*M* = 3452 ms, *SD* = 2787), $F(1, 978) = 17.43$, $p < .001$.

These findings reveal a small difference between the ANOVA and LME analyses. For the LME analyses, there were reliable depletion effects in all three metaphor familiarity groups, whereas in the ANOVAs there was not a reliable depletion effect in the high familiarity metaphor group. Note, however, that the overall pattern was the same--the magnitude of the depletion effect was greater in Low- and Mid-Familiarity metaphors than in High-Familiarity metaphors. The ANOVA and LME analysis both showed significant main effects of Depletion Condition and Stimulus Group.

A summary of the ANOVA and LME statistical results of Experiment 1 is presented in Table II, Appendix A. A summary of subject means and standard deviations from Experiment 1 is presented in Table III, Appendix A.

B. **Experiment 1 - Discussion**

For Experiment 1, I predicted a main effect of familiarity (Hypothesis 1; slower comprehension times as familiarity decreases) and a main effect of depletion condition (Hypothesis 2; slower comprehension times in Depletion than Control condition). More importantly, I predicted a metaphor Stimulus Group X Depletion Condition interaction (Hypothesis 3) in that resource depletion would not significantly impair comprehension times of metaphors that were highly familiar to a reader, but would cause an increasing degree of impairment as metaphors became less familiar to the reader. Results from the ANOVA and LME analyses support these hypotheses. All analyses indicate main effects of Stimulus Group and Depletion Condition, as well a Stimulus Group X Depletion Condition interaction. Furthermore, these analyses all revealed the same pattern of results. Regarding the main effect of depletion, Depletion condition participants were slower in comprehending stimuli than Control condition participants. Regarding the main effect of Stimulus Group, Literal sentences had the fastest comprehension times, followed in order by High-Familiarity metaphors, then Mid-Familiarity metaphors, and finally Low-Familiarity metaphors which had the slowest comprehension times. Regarding the Stimulus Group X Depletion Condition interaction, there was a small depletion effect for High-Familiarity metaphors (Depletion Condition participants respond slower than Control participants), but there was a large depletion effect for Mid- and Low-Familiarity metaphors.

The results of Experiment 1 have several implications. First, the presence of a main effect of Depletion Condition indicates that resource depletion affects comprehension of metaphors. This finding is important because, prior to this study, resource depletion has not been used to study metaphor comprehension. Regarding the observed interaction, larger depletion effects were found in Mid- and Low- familiarity metaphors compared to High-Familiarity metaphors. This pattern indicates that as metaphor familiarity decreases, the impact of resource depletion on comprehension time increases. Conceptually, this may mean that the underlying process of metaphor comprehension changes in some way as metaphor familiarity changes. Specifically, given that the impact of depletion increases as familiarity decreases, the processes involved in metaphor comprehension may become more cognitively taxing as metaphor familiarity decreases. I present a more detailed discussion of this effect and its implications regarding the Career Of Metaphor and Categorization models in the General Discussion.

To verify whether participants evaluated metaphor familiarity as I expected, I compared my participants' ratings of metaphor familiarity to those established by Katz' (1988) metaphor norms. Each of the 45 metaphors used in this study had three associated measures of familiarity. The first of these measures, Katz Familiarity Group, refers to the three groups (high, medium, and low) of metaphor familiarity based on Katz' (1988) metaphor norms. This is a categorical measure as metaphors are assigned to the high-, mid-, and low-familiarity conditions. The second measure, Katz Familiarity Rating, is the actual numerical familiarity rating of each metaphor from Katz' (1988) norming study on a scale of 1 (least familiar) to 7 (most familiar). The range of familiarity ratings for our metaphors was 1.52 to 6.63. The third measure of familiarity, Participant Familiarity Rating, is the average numerical familiarity rating of each metaphor based on ratings given by participants in Experiment 1, again on the 1 to 7 scale. Actual range of participant familiarity ratings was 1.51 to 5.26. Katz's ratings and the actual Participant ratings were highly correlated (r = .78), indicating that my participants' ratings of metaphor familiarity were similar to those established by Katz (1988).

To summarize, the results of Experiment 1 matched my predictions. Resource depletion did slow comprehension time of metaphors (a depletion effect) and depletion effects increased as metaphor familiarity decreased. The implications of these findings for the Categorization and Career of Metaphor models are considered in the General Discussion.

A. **Experiment 2 - Results**

As a reminder, the only difference between Experiments 1 and 2 was the decision made. In Experiment 2, participants indicated whether the sentence is concrete or abstract. Literal sentences should be categorized concrete and metaphors should be categorized as abstract. Response time and accuracy (error rates) were measured. Compared to Experiment 1, there was much less within- and between-subject variability in response times in Experiment 2, as well as fewer outliers. As a result, I did not exclude response times based on an upper threshold. As in Experiment 1, response times under 400 milliseconds were excluded from all analyses.

To account for both subject and item variance in response times, I conducted ANOVA analyses of both subject means (F1) and item means (F2). I also conducted a linear mixed-effects model (LME) analysis on response times to combine subject and item variance to confirm the

ANOVA results. In addition, I conducted an unplanned ANOVA analysis on average per-item error rates as a function of Stimulus Group and Depletion Condition.

For the by-subject analysis, I conducted a 2 (Depletion Condition: Control, Depletion) X 4 (Stimulus Group: Literal, High-Familiarity Metaphor, Mid-Familiarity Metaphor, Low-Familiarity Metaphor) mixed-design ANOVA on average per-subject decision times (in ms). For the by-item analysis, the same analysis was conducted with Depletion Condition as the within-item factor and Stimulus Group as the between-item factor.

The ANOVA revealed that the main effect of Depletion Condition was not significant in the by-subject analysis, $F1(1, 78) = .16$, $MS_e = 47212.59$, $p = .69$, $\omega^2 = .00$. However, the main effect of Depletion Condition was significant in the by-items: items had longer response times by 34 ms in the Depletion condition (*M* = 1569 ms, *SD* = 299) than in the Control condition (*M* = 1535 ms, *SD* = 284), $F2(1, 86) = 4.08$, $MS_e = 10449.79$, $p = .046$, $\omega^2 = .02$. The ANOVA also revealed a main effect of Stimulus Group, which was significant by-subject, $F1(3, 234) = 62.80$, $MS_e = 49001.02$, $p < .001$, $\omega^2 = .45$, and by-items, $F2(3,86) = 26.19$, $MS_e = 86211.49$, $p < .001$, $\omega^2 = .48$. Pairwise comparisons based on data from the by-subject analysis, ordered from fastest to slowest response times, revealed that literal sentences (*M* = 1360 ms, *SD* = 307) had faster response times by 352 ms than Low-Familiarity metaphors (*M* = 1712 ms, *SD* = 431), $p < .001$, faster response times by 357 ms than Mid-Familiarity metaphors (*M* = 1717 ms, *SD* = 470), $p < .001$, and faster response times by 442 ms than High-Familiarity metaphors (*M* = 1802 ms, *SD* = 480), $p < .001$. No significant differences were present in pairwise comparisons among metaphor stimulus groups. Low-Familiarity and Mid-Familiarity metaphors did not differ in response times, $p = .95$. Mid-Familiarity and High-Familiarity metaphors also did not differ in response times, $p = .27$. Finally, Low-Familiarity and High-Familiarity metaphors did not differ in response times, $p = .24$.

No Stimulus Group X Depletion Condition interaction was found by-subject, $F1(3, 234) = .77$, $MS_e = 49001.02$, $p = .51$, $\omega^2 = .01$, or by-items, $F2(3, 86) = 1.35$, $MS_e = 10449.79$, $p = .26$, $\omega^2 = .05$. In the by-subject analysis, observed power was 1.00 for the main effect of Stimulus Group, .07 for the main effect of Depletion Condition, and .21 for the interaction. In the by-item analysis, observed power was 1.00 for the main effect of Stimulus Group, .52 for the main effect of Depletion Condition, and .35 for their interaction.

Because the interaction I hypothesized referred specifically to the different levels of metaphor stimuli familiarity, I also performed two ANOVAs (one by-subject, one by-item) in which only the three metaphor stimulus groups were included (the literal stimuli were excluded). I performed these analyses primarily to see whether a significant Stimulus Group X Depletion Condition interaction would be present when analyzing only the metaphor stimuli. With only the three metaphor stimuli groups included, there was no Stimulus Group X Depletion Condition interaction in either the by-subjects analysis, $F1(2, 156) = 1.27$, $MS_e = 44678.11$, $p = .29$, $\omega^2 = .02$, or the by-item analysis, $F2(2, 42) = 2.06$, $MS_e = 10310.99$, $p = .14$, $\omega^2 = .09$. There was a main effect of Stimulus Group by-subjects, $F1(2, 156) = 4.52$, $MS_e = 44678.11$, $p = .01$, $\omega^2 = .06$, but no main effect of Stimulus Group by-items, $F2(2, 42) = 1.02$, $MS_e = 74201.93$, $p = .37$, $\omega^2 = .05$. Follow-up tests of the significant main effect of Stimulus Group (by-subjects) are not reported here as the pattern of results is the same as in the ANOVA analysis with all four stimulus groups included. Finally, there was no main effect of Depletion Condition in the by-subjects, $F(11, 78) = .13$, $MS_e = 552934.80$, $p = .72$, $\omega^2 = .00$, or by-items, $F2(1, 42) = 2.53$, $MS_e = 10310.99$, $p = .12$, $\omega^2 = .06$.

Linear Mixed-Effect Model Analysis

In the ANOVA results with all four stimulus groups included, the main effect of Depletion Condition was significant in the by-subjects analysis but not the by-items analysis. To resolve this discrepancy, I conducted a linear mixed effect model (LME) analysis (trials nested within individual participants) on comprehension times as a function of Depletion Condition (Control, Depletion) and Stimulus Group (Literal, High-Familiarity, Mid-Familiarity, Low-Familiarity). To match the ANOVA analyses, I performed two LME analyses: one with all four stimulus groups included and one with only the three metaphor groups included. The LME analysis with all four stimulus groups directly matches the by-subjects and by-items ANOVAs which yielded discrepant findings.

With all four stimulus groups included, the LME analysis revealed no main effect of Depletion Condition, $F(1, 3258) = 2.57$, $p = .11$. As with the ANOVAs, there was a main effect of Stimulus Group, $F(3, 2289) = 254.09$, $p < .001$. Pairwise comparisons, presented from fastest to slowest response times, revealed that response times for Literal sentences (M = 1360 ms, SD = 722) were significantly faster by 352 ms than Low-Familiarity metaphors (M = 1712 ms, SD = 792), $p < .001$. Response times for Low-Familiarity and Mid-Familiarity metaphors (M = 1717 ms, SD = 880) did not differ significantly, $p = .70$. Response times for Mid-Familiarity and High-Familiarity metaphors (M = 1802 ms, SD = 916) did not differ significantly, $p = .27$. Finally, as with the ANOVAs, the Depletion Condition X Stimulus Group interaction was not significant, $F(3, 2289) = 1.43$, $p = .23$.

With only the three metaphor stimulus groups included, the LME analysis revealed no main effect of Depletion Condition, $F(1, 2931) = 2.15$, $p = .14$, no main effect of Stimulus Group, $F(2, 2096) = .522$, $p = .59$, and no Depletion Condition X Stimulus Group interaction, $F(2, 2096) = 1.82$, $p = .16$.

Error Rate Analysis

To investigate apparent differences in error rates among the four stimulus groups, I conducted a 2 (Depletion Condition: Control, Depletion) X 4 (Stimulus Group: Literal, High-Familiarity metaphor, Mid-Familiarity metaphor, Low-Familiarity metaphor) mixed design

ANOVA to examine error rates as a function of Depletion Condition (between-subjects) and Stimulus Group (within-subjects).

The ANOVA revealed a main effect of Stimulus Group, $F(3,228) = 27.73$, $MS_e = .01$, $p < .001$, $\omega^2 = .27$, but no main effect of Depletion Condition, $F(1, 76) = .27$, $MS_e = .04$, $p = .61$, $\omega^2 = .00$, and no Stimulus Group X Depletion Condition interaction, $F(3, 228) = .16$, $MS_e = .01$, $p = .92$, $\omega^2 = .00$. I conducted pairwise comparisons among the four stimulus groups to follow up the main effect of Stimulus Group. In order from smallest to largest error rates, pairwise comparisons indicated that Literal sentences ($M = .052$, $SD = .064$) had smaller error rates than Low-Familiarity metaphors ($M = .089$, $SD = .133$), $p = .03$. Low-Familiarity metaphors had error rates that were not significantly different from Mid-Familiarity metaphors ($M = .105$, $SD = .133$), $p = .12$. Finally, Mid-Familiarity metaphors had significantly smaller error rates than High-Familiarity metaphors ($M = .194$, $SD = .179$), $p < .001$.

Because of the variation in error rates among the four stimulus groups, I performed a by-subject ANOVA analysis on response times for trials on which the participant responded correctly only. Given the variation in error rates, it was possible that this analysis would show a different pattern of results compared to analyzing all trials (correct and incorrect). With only correct trials included, the ANOVA revealed no main effect of Depletion Condition, $F(1,76) = .001$, $p = .97$. The ANOVA revealed a main effect of Stimulus Group, $F(3, 228) = 72.81$, $p < .001$. Literal sentences ($M = 1321$ ms, $SD = 281$) had faster response times by 362 ms than Mid-Familiarity metaphors ($M = 1683$ ms, $SD = 417$), $p < .001$, faster response times by 373 ms than Low-Familiarity Metaphors ($M = 1693$ ms, $SD = 390$), $p < .001$, and faster response times by 419 ms than High-Familiarity Metaphors ($M = 1739$ ms, $SD = 422$), $p < .001$. Although no differences among the metaphor stimulus groups met the $p < .05$ significance threshold, the 56 ms difference between High-Familiarity and Mid-Familiarity metaphors was marginally significant ($p = .052$).

Finally, the ANOVA revealed no Depletion Condition x Stimulus Group interaction, $F(3, 228) = .64$, $p = .59$.

A summary of the ANOVA and LME statistical results of Experiment 2 is presented in Table II, Appendix A. A summary of subject means and standard deviations from Experiment 2 is presented in Table IV, Appendix A.

A. Experiment 2 - Discussion

In Experiment 2, my goal was to create a task which would require exertion of cognitive control but would not require participants to fully access the meaning of the presented sentences in order to respond. The abstract/concrete categorization task, which was first used in Pilot Study 3 (see Appendix C), was created to fit these constraints. I reasoned that asking participants to make a concrete/abstract judgment would require them to partially, but not fully, access the meaning of each sentence. By extension, I reasoned that resource depletion would affect overall performance, but that the impact would not vary based on metaphor familiarity. This finding would complement the results of Experiment 1 and allow greater specificity in describing how resource depletion affects processing of metaphor and literal sentences. By using the same metaphor stimuli across Experiments 1 and 2, I could also eliminate the possibility that any differences in depletion effect patterns between the two experiments were caused by using different stimuli.

I expected response times in the Depletion condition to be slower than the Control condition, but the amount of slowing should not vary based on metaphor familiarity. As such, I predicted a main effect of Stimulus Group (Hypothesis 1; longer response times for metaphor than literal stimuli), a main effect of Depletion Condition (Hypothesis 2; longer response times for depletion than control participants) and no Stimulus Group X Condition interaction (Hypothesis 3).

Of my three main predictions, results from both ANOVA and mixed-model analyses supported only Hypothesis 1, as longer response times were observed on metaphor than on literal stimuli. I found no main effect of Depletion Condition, as the difference in response times between the Control and Depletion conditions was minimal for all stimulus groups. Because the main effect of Depletion Condition was a key prediction for Experiment 2, the first step to understanding these results is to consider why the resource depletion manipulation had no impact on response times.

Experiment 2 relied on the assumption that performance of the abstract/concrete categorization task drew on higher-order controlled processes, and would thus be sensitive to the resource depletion induced by the Stroop task. Given that the Stroop task has been used in a number of resource depletion studies and successfully induced depletion in Experiment 1 (observed as a main effect of Depletion Condition), it is unlikely that this task failed to induce depletion in Experiment 2. My assumptions regarding the abstract/concrete categorization task, however, were based largely on the results of Pilot Study 3, which used a limited sample size of 20 participants. Given that this task has not previously been used in research using resource depletion manipulations, the most likely explanation for the observed results is simply that my assumptions about the processing demands of the abstract/concrete categorization task were incorrect. Specifically, categorizing metaphors and literal sentences as “abstract” or “concrete” may not require higher-order controlled processing, at least to a degree which would be significantly affected by depletion.

With the abstract/concrete categorization task, my aim was to create a task which required the same type of effortful, controlled processing from the participant (so as to be sensitive to resource depletion) for all metaphors, regardless of familiarity. I reasoned that performance of this task would require participants to broadly consider the meaning of each sentence without having to determine a precise meaning for each. However, it appears that participants may have used strategies that placed less strain on cognitive resources. For instance, it is possible that participants used ease of comprehension as a proxy for the decision of whether the meaning of each sentence was literal or figurative. With this strategy, sentences for which the meaning was understood automatically would be classified as concrete, whereas those whose meaning was not readily understood were classified as abstract. The use of this or a similar approach would explain why task performance was unimpaired by resource depletion.

Moving beyond my predicted results, an analysis of the error rates in Experiment 2 yields potentially informative results. In the abstract/concrete categorization task, error trials were trials in which the participant categorized a literal sentence as having abstract meaning or a metaphor sentence as having concrete meaning. Error rates, which were expected to be relatively similar across stimulus groups, were significantly larger for high-familiarity metaphors than any other stimulus group. High-familiarity metaphors were incorrectly categorized as concrete more than 20% of the time, compared to roughly 11% for mid- -familiarity metaphors and 9% for low-familiarity metaphors. Literal statements were incorrectly categorized as abstract only 5% of the time.

This pattern of error rates has several potential implications. First, the meaning of literal statements was available very quickly and automatically, which led to easily rating a statement as concrete. Second, the relatively high error rate on highly familiar metaphors may reflect greater automaticity in comprehending these metaphors. If highly familiar metaphors were understood automatically, participants would access their meanings as quickly and easily as those of literal sentences. This ease of comprehension may have led participants to experience highly familiar metaphors as more similar to literal sentences than to less familiar metaphors, leading to categorization errors (responding concrete instead of abstract). In addition, the figurative meaning may be the dominant meaning. This may have led participants to quickly and easily access the figurative meaning, which would lead to an incorrect "concrete" response that had to be suppressed to correctly answer "abstract." This would lead to high error rates.

Analysis of response times in Experiment 2 for correct responses only showed no significant differences in response times among the three metaphor familiarity groups. This result supports the idea that the abstract/concrete categorization task not require participants to fully access the meaning of the presented metaphors, thus nullifying differences in processing difficulty among the three metaphor groups. The idea that the abstract/concrete categorization task did not require full comprehension of the presented metaphors is further supported by comparing response times between Experiments 1 and 2 for the same stimuli. Not only were response times much faster for Experiment 2 compared to Experiment 1, Experiment 2 also shows a much smaller range of response times amongst the three metaphor familiarity groups compared to Experiment 1. Taken together, these data strongly indicate that participants in Experiment 2 did not fully access the meaning of each sentence before responding.

Experiment 2 was intended to complement the findings of Experiment 1. However, my manipulation in Experiment 2 did not work as intended, which might reflect mistaken assumptions regarding the processing demands of the abstract/concrete categorization task. As a result, my goal of using the same stimuli across both experiments and showing that the depletion effect varied as a function of metaphor familiarity in Experiment 1, but not in Experiment 2, was not successful. However, the error rates in Experiment 2 provide support for the idea that highly familiar metaphors are understood automatically. I will discuss this implication further in the General Discussion.

IV. GENERAL DISCUSSION

Experiments 1 and 2 were designed as a complementary pair of experiments that used a resource depletion manipulation to investigate the processing involved in comprehending metaphors of varying levels of familiarity. In Experiment 1, I found that resource depletion impacted metaphor comprehension time differently depending on their familiarity: the depletion effect decreased as metaphor familiarity increased. In Experiment 2, the primary manipulation was unsuccessful. The abstract/concrete decision task was intended to be sensitive to depletion without requiring complete comprehension of the metaphor sentences. Despite the failure of the primary manipulation, analysis of the error rates showed significantly larger error rates for highly familiar metaphors compared to less familiar metaphors. This finding may suggest that highly familiar metaphors are understood automatically.

I set out to compare the predictions of Bowdle and Gentner's (2005) Career of Metaphor (COM) model and Glucksberg's (2003) Categorization model. The two models differ significantly with respect to their descriptions of comprehension of highly familiar metaphors. According to the COM model, highly familiar metaphors are understood automatically, whereas less familiar metaphor comprehension relies on controlled processes that require more time and effort to understand as metaphors become less familiar. By contrast, the Categorization model specifies that the same type of controlled process underlies comprehension of all metaphors, though processing demands increase as familiarity decreases. So, which model do the results support? To begin to answer this question, it is necessary to consider how my findings compare to the ideal results predicted by each model.

To compare my results to the predictions of the Categorization and COM models, it is first necessary to consider that the two models predict a very similar pattern of results for mid- and low-familiarity metaphors. The key difference is the pattern of results on familiar metaphors – the

COM model specifies that they are understood automatically, whereas according to the Categorization model they are understood through the same process that underlies comprehension of less-familiar metaphors. It seems logical to differentiate between the models by examining the Depletion/Control condition time difference for high-familiarity metaphors – if resource depleted participants are slower than controls, the Categorization model is supported; if not, the COM model is supported. However, individual variation among participants prevents such a straightforward analysis. Each participant is not likely to be highly familiar with every one of the 15 metaphors in the high-familiarity set. This variation was reflected in participants' ratings: on a scale from 1 (least familiar) to 7 (most familiar), the average rating for metaphors in the high-familiarity set was 4.33. Thus, even if highly-familiar metaphors are understood automatically, the control/depletion time difference would not be zero for any of the metaphors in the high-familiarity set because none are highly familiar to all participants. Because an observed non-zero control/depletion time difference on high-familiarity metaphors is thus compatible with either the Categorization or COM models, I used a different approach to differentiate between the models.

The strongest piece of evidence differentiating the two models comes from considering the observed differences in response times for each metaphor when used in the Depletion versus the Control conditions, compared to the pattern predicted by the Categorization and COM models. To visualize the ideal findings corresponding to each model's predictions, I considered how the size of the Depletion/Control condition time difference (hereafter referred to as the "depletion effect") for each metaphor compared to its familiarity. Results that aligned perfectly with the Categorization model would show a generally linear increase in the size of the depletion effect as metaphor familiarity decreased, from a small depletion effect for highly familiar metaphors to a large depletion effect for highly unfamiliar metaphors. This corresponds to the model's prediction that metaphor comprehension is a controlled process that becomes more difficult as familiarity decreases. By contrast, results that aligned perfectly with the COM model would show a depletion effect of zero for highly familiar metaphors, and a generally linear increase in size of the depletion effect as familiarity decreased for mid- and low-familiarity metaphors. This break between high-familiarity metaphors and other metaphors corresponds to the COM model's description of separate processes: an automatic process is used to comprehend highly familiar metaphors, whereas a controlled process, which becomes more cognitively taxing as familiarity decreases, is used to comprehend metaphors that are not highly familiar. Figures 1-4 in Appendix A show ideal results aligning with the Categorization (Figure 1) and COM (Figure 2) models, as well as the results showing size of depletion effect as a function of Normed Familiarity (Figure 3) and Rated Familiarity (Figure 4).

In Figures 3 and 4, the trendlines for high-, mid-, and low-familiarity metaphors may be informative in matching the results to the Categorization or COM model. For any given metaphor group, these trendlines indicate whether the size of the depletion effect increases, decreases, or remains constant as familiarity decreases. By extension, this indicates whether, within each metaphor group, comprehension becomes more cognitively taxing as familiarity decreases. In the ideal results aligned with the Categorization model (Figure 1), the trendline has a constant negative slope across all metaphor familiarity groups, indicating that the depletion effect, and thus the cognitive difficulty of comprehending the metaphor, increases as familiarity decreases. In the ideal results for the COM model (Figure 2), the trendlines for low- and mid-familiarity metaphors have a constant negative slope, but the trendline for high-familiarity metaphors has a slope of zero (obscured by the x-axis), indicating that the depletion effect, and thus cognitive difficulty of comprehending the metaphor, does not vary with familiarity within highly familiar metaphors.

In the observed results, the trendlines more closely match the predictions of the COM model than the Categorization model. In Figure 3 (depletion effect as a function of Normed Familiarity) and Figure 4 (depletion effect as a function of Average Rated Familiarity), the trendlines for low- and mid- familiarity metaphors are negatively sloped, whereas the trendlines for high-familiarity metaphors are flat or have a slight positive slope. Specifically, in the results as a function of Normed Familiarity, for every unit decrease in familiarity the depletion effect increases by 383 ms in Low-Familiarity metaphors and by 1661 ms in Mid-Familiarity metaphors, but in High-Familiarity metaphors the depletion effect decreases by 148 ms for every unit decrease in familiarity. Similarly, in the results as a function of Average Rated Familiarity, for every unit decrease in familiarity the depletion effect increases by 1124 ms in Low-Familiarity metaphors and by 812 ms in Mid-Familiarity metaphors, but the depletion effect decreases by 15 ms for every unit decrease in familiarity. This pattern is reflected in the relevant correlations: in low-familiarity metaphors, the depletion effect was negatively correlated with Normed Familiarity (r = -.12) and Average Rated Familiarity (r = -.50). In mid-familiarity metaphors, the depletion effect was also negatively correlated with Normed Familiarity (r = -.47) and Average Rated Familiarity (r = -.35). In contrast, in high-familiarity metaphors, the depletion effect was positively correlated with Normed Familiarity (r = .27) and was virtually uncorrelated with Average Rated Familiarity (r = .02). This indicates a clear break between the high-familiarity metaphor group and the mid- and low-familiarity groups, and suggests that within the high-familiarity metaphors, comprehension difficulty did not increase as familiarity decreased. This matches the predictions of the COM model, but not the Categorization model, which would predict no break between high-familiarity metaphors and mid- and low-familiarity metaphors, but rather constant increase in depletion effect as familiarity decreases within all metaphor groups.

In addition to these trends, considering the error patterns observed in Experiment 2 may be informative. In particular, the unexpectedly large error rate on High-Familiarity metaphors (19.4%), which was approximately twice as large as for Mid-Familiarity metaphors (10.5%) and Low-Familiarity metaphors (9%), may be a reflection of the processes involved in categorizing High-Familiarity metaphors as abstract or concrete. In the discussion for Experiment 2, I speculated that participants used ease and speed of comprehension as a proxy for categorizing a given sentence as concrete in meaning. This explanation would explain both the high error rate on High-Familiarity metaphors and the low error rate on Literal sentences. If highly familiar metaphors are understood automatically, their meanings would be quickly and easily accessible, leading participants to (incorrectly) categorize them as concrete in meaning, whereas the same process would lead participants to (correctly) categorize literal sentences as concrete. Considered alongside the trends discussed in the previous paragraphs, this finding may support the idea that highly-familiar metaphors are understood automatically and thus lend support to the COM model.

Although evidence from depletion effect trends and error rate patterns appears to support the Career of Metaphor model, my results are not wholly incompatible with the Categorization model. The overall results pattern of increasing comprehension times and increasing depletion effect as familiarity decreases is compatible with the predictions of both models. The key in differentiating the two models is whether this overall pattern is interpreted as representing a categorical shift from automatic to controlled processing or a spectrum of low to high effort controlled processes. Although I have described several pieces of evidence that favor the first interpretation, some individual results favor the second. For instance, results from the LME analyses in Experiment 1indicated that resource depletion led to slower comprehension on all metaphor groups, including high-familiarity metaphors. Although the existence of a depletion effect on some high-familiarity metaphors may be attributed to individual variability in familiarity with the presented metaphors, it may be taken as indicating that comprehension of some familiar metaphors is a controlled process. This interpretation would support the Categorization model's description of a controlled process used to comprehend metaphors of any degree of familiarity.

The Career of Metaphor and Categorization models make predictions that are difficult or impossible to differentiate by simply measuring comprehension times of metaphors of different levels of familiarity. As such, this study aimed to differentiate between the two models by using a methodology that would be sensitive to the differences in the underlying processes of metaphor comprehension that each model predicts. In particular, because resource depletion affects only higher-order controlled processing, I reasoned that a resource depletion methodology could be used to differentiate between the two models. However, due to variability in different individuals' familiarity with any given metaphor, ambiguities remain in separating the two models. Nevertheless, based on the logic outlined in the preceding paragraphs, the results of this study may lend tentative support to the Career of Metaphor model over the Categorization model.

This study suggests several methodological and conceptual directions for future research. First, regarding methodology, this study shows that resource depletion is sensitive to processing differences between metaphors of different levels of familiarity. Thus, future research on metaphors or other figurative language may consider using a resource depletion manipulation in similar contexts, such as when studying different types of phrases or texts which are believed to elicit different levels of cognitive exertion. On a conceptual level, the results of this study may inform future research investigating whether the most highly-familiar metaphors are processed differently than other metaphors. For instance, a question posed by the present study is, if the COM model's description of separate processes is correct, where is the threshold between automatic and controlled processing? For instance, how many times does one need to be exposed to a metaphor before comprehension becomes automatic? If a given metaphor is highly familiar, does it need to be presented in the exact surface form that is familiar to the reader in order to be understood automatically, or is this automaticity resilient to changes in surface form? One possibility is that only the metaphors that are most familiar to an individual – perhaps ones which have a standardized surface form and are commonly encountered in everyday discourse – are comprehended automatically. However, my results indicate that the size of the depletion effect, and by extension the degree of cognitive exertion, was constant across the high-familiarity stimuli, which spanned a range from very highly familiar to only moderately highly familiar. This finding may suggest that even metaphors which an individual is only somewhat highly familiar with are understood automatically. Future research should investigate whether automatic comprehension is a phenomenon that applies to only a small number of commonly encountered metaphors, or whether a significant percentage of metaphors are understood this way.

REFERENCES

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: is the active self a limited resource? *Journal of personality and social psychology*, *74*(5), 1252.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current directions in psychological science*, *16*(6), 351-355.

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological review*, *112*(1), 193.

Bray, S. R., Martin Ginis, K. A., Hicks, A. L., & Woodgate, J. (2008). Effects of self-regulatory strength depletion on muscular performance and EMG activation. *Psychophysiology*, *45*(2), 337-343.

Campbell, S. (2014). *Evaluating the Career of Metaphor and Categorization Models of Metaphor Processing: An Eye--Tracking Study* (Doctoral dissertation, University of Illinois at Chicago).

Campbell, S. J., & Raney, G. E. (2015). A 25-year replication of Katz et al.'s (1988) metaphor norms. *Behavior research methods*, 1-11.

Coney, J., & Lange, A. (2006). Automatic and attentional processes in the comprehension of unfamiliar metaphors. *Current Psychology*, *25*(2), 93-119.

Gentner, D., & Bowdle, B. (2002). Psychology of metaphor processing.*Encyclopedia of Cognitive Science. London: Nature Publishing Group*.

Gibbs Jr, R. W., & Gerrig, R. J. (1989). How context makes metaphor comprehension seem'special'. *Metaphor and Symbol*, *4*(3), 145-158.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of personality and social psychology*, *54*(5), 733.

Gildea, P., & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 577-590.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in cognitive sciences*, *7*(2), 92-96.

Glucksberg, S. (2008). How metaphors create categories–quickly. *The Cambridge handbook of metaphor and thought*, 67-83.

Glucksberg, S., Gildea, P., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of verbal learning and verbal behavior*, *21*(1), 85-98.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: a meta-analysis. *Psychological bulletin*, *136*(4), 495.

Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbol*, *3*(4), 191-214.

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, *1284*, 145-155.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive science*, *4*(2), 195-208.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, *109*(2), 163.

Minkoff, S. R., & Raney, G. E. (2000). Letter-detection errors in the word the: Word frequency versus syntactic structure. *Scientific Studies of Reading*,*4*(1), 55-76.

Oaten, M., & Cheng, K. (2006). Improved self-control: The benefits of a regular program of academic study. *Basic and Applied Social Psychology*,*28*(1), 1-16.

Oaten, M., & Cheng, K. (2006). Longitudinal gains in self-regulation from regular physical exercise. *British journal of health psychology*, *11*(4), 717-733.

Oaten, M., & Cheng, K. (2007). Improvements in self-control from financial monitoring. *Journal of Economic Psychology*, *28*(4), 487-501.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay effects of interracial contact on executive function. *Psychological Science*, *14*(3), 287-290.

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: role of the self in logical reasoning and other information processing. *Journal of personality and social psychology*, *85*(1), 33.

Searle, J. R. (1975). The logical status of fictional discourse. *New literary history*, *6*(2), 319-332.

Smit, A. S., Eling, P. A., & Coenen, A. M. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta psychologica*, *115*(1), 35-42.

Therriault, D. J., & Raney, G. E. (2007). Processing and representing temporal information in narrative text. *Discourse Processes*, *43*(2), 173-200.

Vohs, K. D., & Heatherton, T. F. (2000). Self-regulatory failure: A resource-depletion approach. *Psychological science*, *11*(3), 249-254.

Vohs, K. D., Baumeister, R. F., & Ciarocco, N. J. (2005). Self-regulation and self-presentation: regulatory resource depletion impairs impression management and effortful self-presentation depletes regulatory resources.*Journal of personality and social psychology*, *88*(4), 632.

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology, 94(5)*, 883–898.

Webb, T. L., & Sheeran, P. (2003). Can implementation intentions help to overcome ego-depletion? *Journal of Experimental Social Psychology*, *39*(3), 279-286.

Appendix A

Tables and Figures

TABLE I

SAMPLE SENTENCE STIMULI FOR EXPERIMENTS 1 AND 2

Literal Comparison	High-Familiarity Metaphor	Mid-Familiarity Metaphor	Low-Familiarity Metaphor
A skyscraper is a building	The mind is a sponge	Sleep is an ocean	A fisherman is a spider
A cucumber is a vegetable	Love is a flower	A smile is an ambassador	Creativity is a toaster
A hammer is a tool	My roommate is a pig	An accountant is a juggler	Truth is a firefly
A flamingo is a bird	Alcohol is a crutch	Danger is a spice	Wisdom is a foreigner

TABLE II

SUMMARY OF RESULTS OF ANOVA AND LME ANALYSES IN EXPERIMENTS 1 AND 2

Experiment	**Analysis**	**Stimulus Groups Included**	**Analysis Type**	**Effect DC = Depletion Condition SG = Stimulus Group**	***F***	***p***	**ω^2**
Experiment 1	ANOVA	All Four Stimulus Groups	By Subject	Main Effect of DC	4.03	.048	.05
				Main Effect of SG	111.42	<.001	.59
				DC x SG Interaction	3.18	.025	.04
			By Item	Main Effect of DC	91.15	< .001	.52
				Main Effect of SG	187.42	< .001	.87
				DC x SG Interaction	14.36	< .001	.33
		Metaphor Stimulus Groups Only	By Subject	Main Effect of DC	4.06	.047	.05
				Main Effect of SG	79.55	< .001	.50
				DC x SG Interaction	3.31	.04	.04
			By Item	Main Effect of DC	50.44	< .001	.55
				Main Effect of SG	38.36	< .001	.65
				DC x SG Interaction	4.18	.02	.17
	LME	Metaphor Stimulus Groups Only	N/A	Main Effect of DC	42.47	<.001	N/A
				Main Effect of SG	112.71	<.001	N/A
				DC x SG Interaction	4.21	.02	N/A
Experiment 2	ANOVA	All Four Stimulus Groups	By Subject	Main Effect of DC	0.16	.69	.00
				Main Effect of SG	62.80	< .001	.45
				DC x SG Interaction	0.77	.51	.01
			By Item	Main Effect of DC	4.08	.046	.04
				Main Effect of SG	26.19	<.001	.48
				DC x SG Interaction	1.35	.26	.05
		Metaphor Stimulus Groups Only	By Subject	Main Effect of DC	0.13	.72	.00
				Main Effect of SG	4.52	.01	.06
				DC x SG Interaction	1.27	.29	.02
			By Item	Main Effect of DC	2.53	.12	.06
				Main Effect of SG	1.02	.37	.05
				DC x SG Interaction	2.06	.14	.09
	LME	All Four Stimulus Groups	N/A	Main Effect of DC	2.57	.11	N/A
				Main Effect of SG	254.09	< .001	N/A
				DC x SG Interaction	1.43	.23	N/A

TABLE III

SUMMARY OF SUBJECT MEANS AND STANDARD DEVIATIONS IN EXPERIMENT 1.

All values are in milliseconds; values in parentheses indicate standard deviations.

		Stimulus Group				
		Literal	High-Fam Metaphor	Mid-Fam Metaphor	Low-Fam Metaphor	***M***
Depletion Condition	Control	1517 (564)	2276 (1004)	2861 (1350)	3445 (1736)	2425 (1423)
	Depletion	1634 (638)	2514 (857)	3595 (1603)	4201(1656)	2986 (1601)
	M	1576 (601)	2395 (935)	3228 (1518)	3823 (1728)	

TABLE IV

SUMMARY OF SUBJECT MEANS AND STANDARD DEVIATIONS IN EXPERIMENT 2

All values are in milliseconds; values in parentheses indicate standard deviations.

		Stimulus Group				
		Literal	High-Fam Metaphor	Mid-Fam Metaphor	Low-Fam Metaphor	***M***
Depletion Condition	Control	1342 (272)	1802 (491)	1670 (436)	1709 (448)	1535 (808)
	Depletion	1378 (340)	1802 (475)	1765 (502)	1715 (419)	1569 (831)
	M	1360 (307)	1802 (480)	1717 (469)	1712 (431)	

Figure 1. Ideal results aligning with the Categorization model (Glucksberg, 2003) of metaphor comprehension. Size of depletion effect for literal sentences and high-, mid-, and low-familiarity metaphors as a function of familiarity. The size of the depletion effect (how much slower, on average, Depletion participants respond compared to Control participants for each metaphor) increases in a linear fashion as familiarity decreases. Literal statements, assigned a value of 7 on familiarity, have a depletion effect of zero.

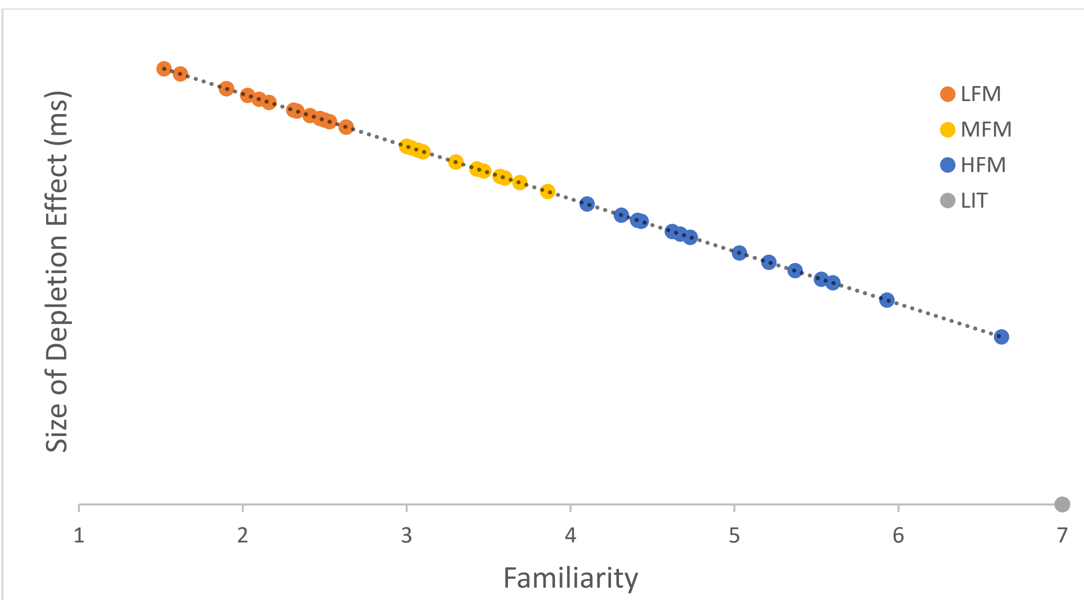


Figure 2. Ideal results aligning with the Career of Metaphor model (Bowdle & Gentner, 2005) of metaphor comprehension. Size of depletion effect for literal sentences and high-, mid-, and low-familiarity metaphors as a function of familiarity. Highly-familiar metaphors, which the COM model predicts are understood automatically, have a depletion effect of zero. For mid- and low-familiarity metaphors, the size of the depletion effect (how much slower, on average, Depletion participants respond compared to Control participants for each metaphor) increases in a linear fashion as familiarity decreases. Literal statements, assigned a value of 7 on familiarity, have a depletion effect of zero.

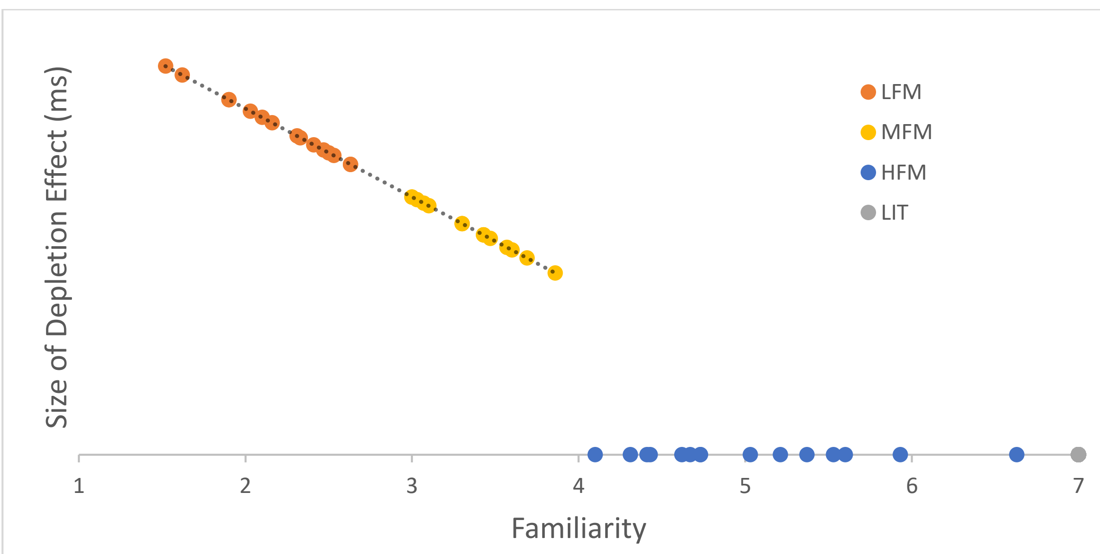


Figure 3. Observed results – size of depletion effect for literal sentences and high-, mid-, and low-familiarity metaphors as a function of Normed Familiarity (from Katz, 1988). Trendlines represent the relationship between size of depletion effect and Normed Familiarity for each metaphor group. Whereas the trend for both mid- and low-familiarity metaphors is an increase in depletion effect as familiarity decreases, the trend for high-familiarity metaphors is a slight decrease in depletion effect as familiarity decreases. Literal statements have been assigned a value of 7 on Normed Familiarity.

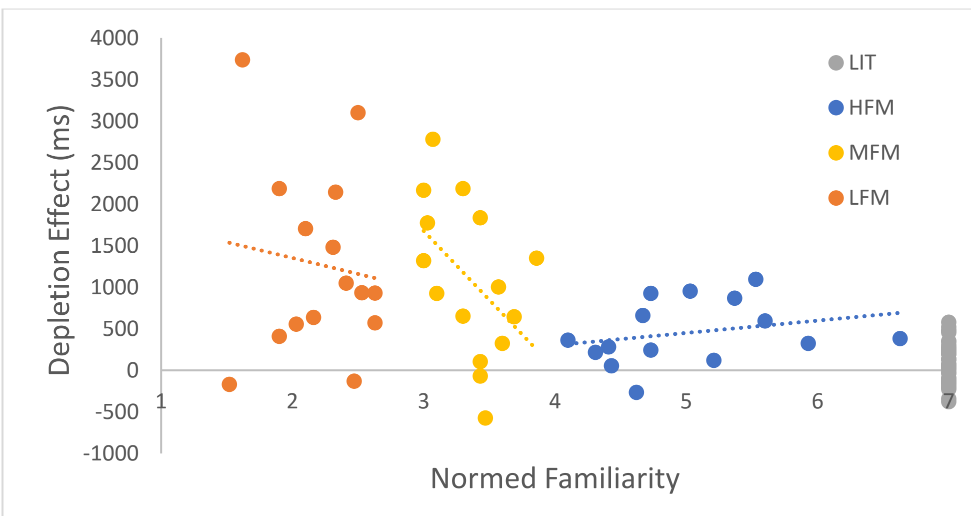
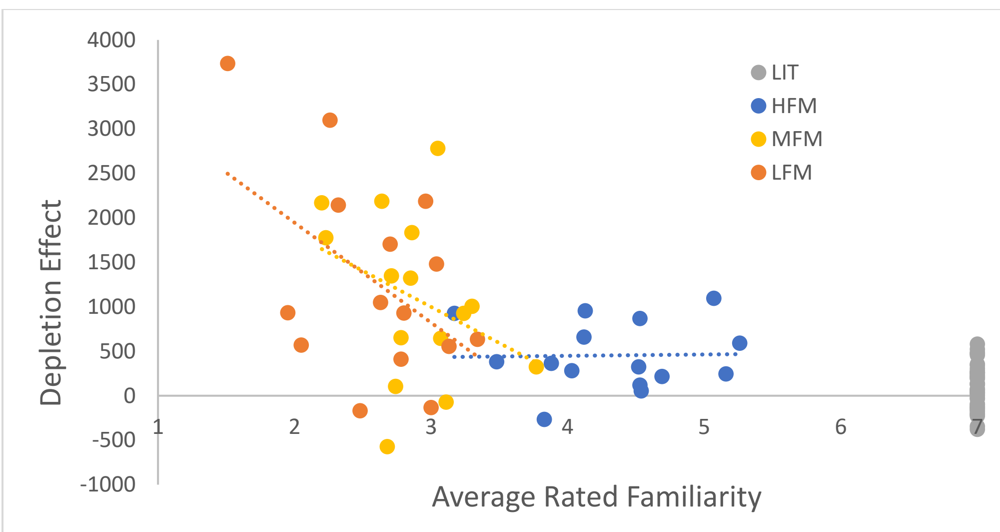


Figure 4. Observed results – size of depletion effect for literal sentences and high-, mid-, and low-familiarity metaphors as a function of Average Rated Familiarity. Trendlines represent the relationship between size of depletion effect and Normed Familiarity for each metaphor group. Whereas the trend for both mid- and low-familiarity metaphors is an increase in depletion effect as familiarity decreases, the trend for high-familiarity metaphors is a slight decrease in depletion effect as familiarity decreases. Literal statements have been assigned a value of 7 on Normed Familiarity.



Appendix B

UIC Vocabulary Quiz and Language History Questionnaire

Vocabulary Test (Version 6/09/2004) **Subject __________**

Directions: Choose the BEST definition for each word.

1. ASCEND
A. to go up or mount
B. consent
C. improve with time
D. to leave behind
E. to replace a leader

2. WARY
A. tired out
B. rude; uncouth
C. perturbed
D. brand-new
E. cautious; careful

3. NURTURE
A. helped by man
B. to feed or nourish
C. to educate
D. to protect by nature
E. to cook

4. INFINITESIMAL
A. very long
B. very slow
C. well defined
D. uncompromising
E. very small

5. BELLIGERENT
A. informative
B. blunt
C. tiring
D. war-like
E. pro-active

6. INDIFFERENT
A. similar
B. unconcerned
C. diffident
D. solicitous
E. opposite

7. PERJURE
A. to save from indignity
B. to improve or rectify
C. to demand support
D. to lie under oath
E. day by day

8. VERBOSE
A. slow
B. impressive
C. complicated
D. wordy
E. meaningless

9. OPAQUE
A. transparent
B. slippery
C. impenetrable by light
D. gem-like
E. financially well-off

10. SYNTHESIS
A. musical rendition of a written work
B. a theory of immoral behavior
C. the combination of parts to form a whole
D. watching or guarding
E. properties of artificial chemicals

11. SPONTANEITY
A. unwanted laughter
B. uncontrollable danger
C. unplanned action
D. unneeded socialism
E. stand-up attitude

12. VALIDATE
A. to prove
B. to get paid back
C. to expire
D. to run away
E. to complete successfully

13. SUBORDINATE
A. to hypothesize in abstract
B. to practice with instruction
C. to levy upon others
D. to go on vacation
E. to rank in importance

14. MEAGER
A. not full, inadequate
B. to beg
C. without self-respect
D. in good shape, healthy
E. wise, full of advice

15. EQUIVOCAL
A. premier, establishing new precedent
B. popular, known by everyone
C. exciting, causing a commotion
D. peculiar, one of a kind
E. uncertain, having two meanings

16. REBUKE
A. to dispute
B. poor reputation
C. to scold harshly
D. to stop at midpoint
E. to overfill

17. ECLECTIC
A. providential
B. of religious origins
C. purified
D. out of fashion
E. from various sources

18. TERSE
A. concise
B. private
C. angry
D. outdated
E. harsh-sounding

19. ILLUSORY
A. bright
B. deceptive
C. unhealthy
D. making a reference to
E. sometimes friendly, sometimes undependable

20. DIVULGE
A. to discourage
B. to pay for
C. to turn away
D. to reveal
E. to infiltrate

21. REPROVE
A. to reverse an argument
B. to be clean of
C. to express disapproval
D. to grovel for forgiveness
E. to encourage hope

22. IMPLAUSIBLE
A. could happen at any moment
B. not believable
C. unyielding
D. considered tactless
E. to serve or worship

23. INCONTROVERTIBLE
A. useless
B. prone to trouble making
C. indisputable
D. successful
E. unprotected

24. QUERY
A. excavation
B. prey
C. inquiry
D. strange occurrence
E. strange, odd

25. DISPERSE
A. to seize one's assets
B. to live in exile
C. to break up and scatter
D. to weaken connections
E. to make vacant

26. VACILLATE
A. to prepare for action; lubricate
B. to show indecision; to waver
C. to hold firmly, to be stubborn
D. to wait until the last second, delay
E. to scatter; to create chaos

27. SUPERFLUOUS
A. gay, happy
B. reserved, waiting
C. trivial; unimportant
D. unnecessary; excessive
E. undecided; variable

28. AUTONOMOUS
A. unknown identity
B. having many names
C. uncontrollable
D. independent existence
E. self-confidence

29. PRECEDENT
A. an expectation
B. most important event
C. a leader
D. a prior occurrence
E. a forgotten time

30. BOLSTER
A. to disagree, strongly
B. to defend, proudly
C. to reinforce, strengthen
D. to agonize, repeatedly
E. brutalize, mercilessly

Language History Questionnaire

Subject # ________ (Version 12-12-2001)

Sex _________ Age _________ What country were you born in? _______________________

Years living in U.S. _________ Years in U.S. Schools _________

(1) What is the FIRST language you spoke? If your parents spoke two languages to you, list BOTH languages.

(2) List from MOST fluent to LEAST fluent all of the languages that you know (write on the back of this page if you need more space). Note that the language you learned first is not necessarily the language you now know best. Specify the age at which you began to learn the language (if it is your native language you should specify age as "birth") and where you learned it (e.g., school, home, church).

	Language	Age learned	Location learned
Most fluent	____________________	__________	______________________________
	____________________	__________	______________________________
	____________________	__________	______________________________
Least fluent	____________________	__________	______________________________

(3) Answer the following questions. Complete only those questions that apply to you.
At what age did you begin <u>speaking</u> **English**? __________
At what age did you begin <u>reading</u> **English**? __________
At what age did you begin <u>speaking</u> your most fluent language **OTHER THAN** English? __________
At what age did you begin <u>reading</u> your most fluent language **OTHER THAN** English? __________

(4) Complete the following ratings. If you think you are more proficient in either English or your OTHER language, your ratings should reflect this difference. Answer only those questions that apply to you.

	NOT fluent									VERY fluent
For ENGLISH:										
How fluent are you in <u>speaking</u>?	1	2	3	4	5	6	7	8	9	10
How fluent are you in <u>understanding</u>?	1	2	3	4	5	6	7	8	9	10
How fluent are you in <u>reading</u>?	1	2	3	4	5	6	7	8	9	10
For your most fluent language OTHER THAN English:										
How fluent are you in <u>speaking</u>?	1	2	3	4	5	6	7	8	9	10
How fluent are you in <u>understanding</u>?	1	2	3	4	5	6	7	8	9	10
How fluent are you in <u>reading</u>?	1	2	3	4	5	6	7	8	9	10

Appendix C

Methodology for Pilot Studies 1, 2, and 3.

Pilot Study 1

Participants

A total of 16 participants were recruited for pilot study 1. Participants were undergraduate students in the University of Illinois at Chicago (UIC) psychology subject pool who volunteered for the study in exchange for course credit. Participants were proficient English speakers, which was defined as attending and English speaking school for at least 10 years. All participants were administered a language history questionnaire and vocabulary quiz.

Stimuli

A total of 60 stimuli were created for Experiment 1. Of these, 15 were sentences made from high-familiarity metaphors, 15 were low-familiarity metaphors, and 30 were literal comparisons constructed to match the structure of a metaphor. Literal comparison sentences were constructed such that the average word length and word frequency of the first and second nouns in the literal sentences was equal to the average word frequencies of the target and vehicle words, respectively, of the metaphor sentences. An example of a literal sentence is *a flamingo is a bird*. Sentences were presented in a pseudorandomized order, manipulated such that no one type of sentence (high-familiarity metaphor, low-familiarity metaphor, or literal) appeared more than three times in a row. A table of sample stimuli is presented in Table I, Appendix A.

Katz's (1988) database contains 256 metaphors normed a set of ten criteria including familiarity. For this experiment, "high-familiarity" metaphors were selected from among metaphors in the top 25% of familiarity ratings and "low-familiarity" metaphors were selected from metaphors in the bottom 25% of familiarity ratings.

Depletion Task

The depletion manipulation was a video task developed by Baumeister and Tice (adapted from Gilbert, Pelham, & Krull, 1988) in which participants watched a six-minute long silent video of a woman being interviewed. The Baumeister and Tice Psychology Lab makes this video freely available on their website so that it may be used by other researchers (https://psy.fsu.edu/~baumeisterticelab/egodepletion.html). In the video, a series of words (not thematically related to the video) are simultaneously presented in the bottom-right corner of the screen for five seconds each. Participants in the depletion condition were instructed to not look at the words and to quickly refocus on the female interviewee if they happen to glance at the words. Participants in the control condition were given no instruction regarding the words. Because the words occupy a relatively large area of the screen and are not easily ignored, and also because reading each word is an automatic process once the word is seen, participants must attempt to suppress the desire to read the words by constantly diverting their gaze in order to successfully follow the depletion instructions. Thus, successfully following the depletion condition instructions requires the exertion of cognitive control, whereas following the control condition instructions does not.

Procedure

Participants performed the experiment in a quiet room in groups of one to three. Participants began by signing an informed consent form. They then completed the UIC English Vocabulary Quiz and the Language History Questionnaire, which were presented via an online Qualtrics survey.

The remainder of the experiment consisted of two tasks: the video depletion manipulation and the metaphor/literal statement categorization task. The experimenter gave a brief verbal explanation of the video task, the instructions for which varied depending on whether the current participant(s) were in the control or depletion condition. In both conditions, the experimenter instructed the participants to “carefully watch the female interviewee’s nonverbal communication and try to get a sense of how she’s doing in the interview, so that you can rate her communication skills later in this study”. In the resource depletion condition, the experimenter additionally instructed the participant(s) that “there will be some words that come up in the bottom-right hand corner of the screen. These words are not related to the video in any way. For this study, it’s very important that you not look at or read the words at all, but just focus on the video of the woman being interviewed. If you catch yourself glancing down at the words, just make sure to refocus on the video right away. It might be difficult, but it’s important for the results that you focus your attention on the video and not look at the words at all.” Participants in the control condition received no instruction regarding the words. The experimenter also answered any questions that participants may have had before starting the video. When participants finished the video task, the experimenter gave verbal instructions for the metaphor and literal sentence comprehension task

In the metaphor and literal sentence comprehension task, participants were presented with 60 metaphor and literal phrases arranged in a pseudo-randomized order, centered on a computer screen. Participants responded using a response key to indicate the moment at which they have understood the presented sentence or arrived at a reasonable interpretation of the sentence. The Superlab program (Cedrus, 2010) controlled all aspects of the experiment, such as timing of stimuli presentation and recording decision times. After each trial, participants were able to advance to the next trial by pressing a key, upon which a fixation cross was presented in the center of the screen for 500 ms followed by presentation of the next phrase. Five practice trials were presented to allow participants to become familiar with the task. After completing the metaphor/literal categorization task, participants were given a written debriefing form describing the study's goals. The duration of the experiment did not exceed one hour.

Pilot Study 2

Participants

A total of 40 participants from the same population as pilot study 1 participated.

Materials and Procedure

Pilot study 2 was identical to pilot study 1 except for the decision made. In pilot study 2, participants indicated whether the statements were "literal" or "figurative" by pressing one of two response keys. Participants were instructed as follows: "In this task you will read a number of short sentences. For each sentence, you will need to decide, as quickly as possible, whether the meaning of that sentence is literal or figurative. If the meaning is literal, press the response button labeled "literal". If the meaning is figurative, press the response button labeled "figurative"."

Pilot Study 3

Participants

A total of 20 participants from the same population as pilot studies 1 and 2 participated.

Materials and Procedure

Pilot study 3 was identical to pilot studies 1 and 2 except for the decision made. In pilot study 3, participants indicated whether the statements were "concrete" or "abstract" by pressing one of two response keys. Participants were instructed as follows: "In this task you will read a number of short sentences. For each sentence, you will need to decide, as quickly as possible, whether the meaning of that sentence is concrete or abstract. The meaning of a sentence is concrete if it is referring to objects or relationships that are physically real. For instance, the meaning of the sentence *a hammer is a tool* is concrete. The meaning of a sentence is abstract if it refers to ideas or relationships that are not physically real. For instance, the meaning of the sentence sentence *love is a flower* is abstract. Respond "concrete" or "abstract" based on the meaning of the sentence as a whole, not on the individual words in the sentence. If the meaning is concrete, press the response button labeled "concrete". If the meaning is abstract, press the response button labeled "abstract"."

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Approval Notice
Initial Review (Response To Modifications)

October 20, 2016

Felix Pambuccian
Psychology
1007 W, Harrison
Psychology, M/C 285
Chicago, IL 60607
Phone: (651) 494-7939

RE: Protocol # 2016-0780
"Metaphor Comprehension and Processing"

Dear Mr. Pambuccian:

Please note that stamped and approved .pdfs of all recruitment and consent documents will be forwarded as an attachment to a separate email. OPRS/IRB no longer issues paper letters and stamped/approved documents, so it will be necessary to retain the emailed documents for your files for auditing purposes.

Kindly remember that only mass testing/eligibility screening instruments that have been identified as such, and approved and stamped by the IRB, may be used in the mass testing sessions conducted at the beginning of each term.

Your Initial Review (Response To Modifications) was reviewed and approved by the Expedited review process on October 20, 2016. You may now begin your research

Please note the following information about your approved research protocol:

Protocol Approval Period: October 20, 2016 - October 20, 2017
Approved Subject Enrollment #: 1000
Additional Determinations for Research Involving Minors: The Board determined that this research satisfies 45CFR46.404 ', research not involving greater than minimal risk. Therefore, in accordance with 45CFR46.408 ', the IRB determined that only one parent's/legal guardian's permission/signature is needed. Wards of the State may not be enrolled unless the IRB grants specific approval and assures inclusion of additional protections in the research required under

45CFR46.409 '. If you wish to enroll Wards of the State contact OPRS and refer to the tip sheet.
Performance Sites: UIC
Sponsor: None
PAF#: - Not applicable
Research Protocol(s):

a) Metaphor Comprehension and Processing; Version 1.2; 10/03/2016

Recruitment Material(s):

a) No recruitment materials will be used-UIC Psychology Subject Pool procedures will be followed.

Informed Consent(s):

a) Metaphor Comprehension (30 minutes); Version 1,1; 09/02/2016
b) Metaphor Comprehension (60 minutes); Version 1.1; 09/02/2016
c) Metaphor Comprehension (90 minutes); Version 1.1; 09/02/2016
d) Debriefing Form; Version 1; 10/03/2016

Parental Permission(s):

a) A waiver of parental permission has been granted for minors in the UIC Psychology Subject Pool under 45 CFR 46.116(d) and 45 CFR 46.408(c); however, as per UIC Psychology Subject Pool policy, as least one parent must sign the Blanket Parental Permission document prior to the minor subject's participation in the UIC Psychology Subject Pool.

Your research meets the criteria for expedited review as defined in 45 CFR 46.110(b)(1) under the following specific category(ies):

(7) Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Please note the Review History of this submission:

Receipt Date	Submission Type	Review Process	Review Date	Review Action
07/28/2016	Initial Review	Expedited	08/14/2016	Modifications Required
09/06/2016	Response To Modifications	Expedited	09/28/2016	Modifications Required
10/04/2016	Response To Modifications	Expedited	10/20/2016	Approved

Please remember to:

→ Use your **research protocol number** (2016-0780) on any documents or correspondence with the IRB concerning your research protocol.

→ Review and comply with all requirements on the OPRS website at,
"**UIC Investigator Responsibilities, Protection of Human Research Subjects**"
(http://tigger.uic.edu/depts/ovcr/research/protocolreview/irb/policies/0924.pdf)

Please note that the UIC IRB has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact OPRS at (312) 996-1711 or me at (312) 355-0816.

Sincerely,

Alison Santiago, MSW, MJ
Assistant Director, IRB # 2
Office for the Protection of Research
Subjects

Enclosure(s) will be sent in a separate email:

1. **Informed Consent Document(s):**
   a) Metaphor Comprehension (30 minutes); Version 1,1; 09/02/2016
   b) Metaphor Comprehension (60 minutes); Version 1.1; 09/02/2016
   c) Metaphor Comprehension (90 minutes); Version 1.1; 09/02/2016
   d) Debriefing Form; Version 1; 10/03/2016

cc: Michael E. Ragozzino, Psychology, M/C 285
Gary Raney (Faculty Advisor), Psychology, M/C 285

CURRICULUM VITAE

Felix S. Pambuccian

292 Bartram Rd
Riverside, IL 60548

Phone: (651) 494-793
Email: fpambu2@uic.edu

EDUCATION

2014-present	University of Illinois at Chicago, Chicago, IL Graduate Student Ph.D., Psychology Concentration: Cognitive Psychology
2014-2017	University of Illinois at Chicago, Chicago, IL M.A., Psychology Concentration: Cognitive
2008-2013	University of Minnesota, Minneapolis, MN B.S., *summa cum laude* Major: Psychology Minor: French

FELLOWSHIPS, AWARDS, AND HONORS

2014-present	University Fellowship, University of Illinois at Chicago
2011-present	Psi Chi
2010-present	Phi Beta Kappa

TEACHING EXPERIENCE

2015-2017	Teaching Assistant, Research Methods in Psychology University of Illinois at Chicago Supervisor: Dr. Eric Gobel
2016-2017	Teaching Assistant, Psychology of Cognition and Memory University of Illinois at Chicago Supervisor: Dr. Christopher Baker
May 2017	Guest Lecture for Psychology of Cognition and Memory class Topic: Language

ACADEMIC PRESENTATIONS

Pambuccian, F.S., & Raney, G.E. (2016, November). Beyond the word level: Using letter detection to investigate reading of metaphor phrases. Poster presented at the 57th annual meeting of the Psychonomic Society, Boston, MA.

Pambuccian, F.S., & Raney, G.E. (2016, May). Using letter detection to investigate reading of metaphor phrases. Poster presented at the Midwestern Psychological Association, Chicago, IL.

DEPARTMENT PRESENTATIONS

Pambuccian, F.S. (2016, October). Comparing processing of familiar and unfamiliar metaphors using a resource depletion manipulation. Oral presentation for University of Illinois at Chicago Cognitive Psychology Current Research Topics Brown Bag.

Pambuccian, F.S. (2016, October). Beyond the word level: Using the letter detection task to investigate metaphor processing. Oral presentation for University of Illinois at Chicago Cognitive Psychology Current Research Topics Brown Bag.